# In Search of a New Language

## Measuring Style of Góngora and Picasso

Nanette Rißler-Pipka (Karlsruhe)

ABSTRACT: Both authors are considered to be 'unreadable', both are separeted by more than 300 years, both authors are said to have invented a new language or to use language in a mannered way. It is well known that Picasso, like many of his contemporaries, was fascinated by Góngora and made broadly use of other artworks and literature. But how can one compare both authors using digital methods? The language of the 16th/17th century cannot be compared with that of the 20th century. Nevertheless, there are ways and means of opening up new perspectives on the subject of author and style. Here are examples of different digital methods presented and used: Part-of-speech-tagging, stylometry with stylo for R and Topic Modeling. In addition, the connection between stylistics and stylometry is discussed.

ZUSAMMENFASSUNG: Beide Autoren hält man für ‚unlesbar', beide trennen mehr als 300 Jahre. Von beiden Autoren sagt man sie hätten eine neue Sprache erfunden oder die alte Sprache verkünstelt. Dass Picasso wie viele seiner Zeitgenossen von Góngora fasziniert war und sich in der Kunst- wie in der Literaturgeschichte bediente, ist bekannt. Wie soll man aber auf Ebene des digitalisierten Werks beide Autoren vergleichen? Die Sprache des 16./17. Jahrhunderts lässt sich nicht mit derjenigen des 20. Jahrhunderts vergleichen. Dennoch gibt es Mittel und Wege, neue Perspektiven auf das Thema Autor und Stil zu eröffnen. Hier werden exemplarisch verschiedene digitale Methoden vorgestellt und verwendet: Part of Speech Tagging, Stilometrie mit stylo für R und Topic Modeling. Außerdem wird die Verbindung zwischen Stilistik und Stilometrie diskutiert.

SCHLAGWÖRTER: Picasso, Pablo; Góngora, Luis de; Autorschaft; Stil; Stilometrie; Topic Modeling

Today, concepts of intertextuality make it possible to compare authors from completely different times and style, because every author takes his ideas, language and some parts of his style from his predecessors. Picasso as a painter is particularly known for recreating masterpieces of art history (Velázquez, El Greco, Goya, Manet, etc.), but he is rather unknown for his poetry, and for reinventing Góngora's style of obscure and dazzling poetry.[1]

[1] Picasso also illustrated Góngora's sonnets in 1948, for an analysis in context with his writings see Nanette Rißler-Pipka, *Picassos schriftstellerisches Werk: Passagen zwischen Bild und Text* (Bielefeld: transcript, 2015), 297–313.

About both poets, Góngora and Picasso, we can read in contemporary and actual literature that they are "unreadable", treating language as a puzzle to be solved or at least as game of combinatorics (trying new syntaxes, new words, new grammatical orders). Speaking about combinatorics, there are many points that link mathematics, digital humanities and literary criticism. Borges comments on the importance of combinatorics, games and literature:[2]

> hacer de la metafísica, y de las artes, una suerte de juego combinatorio. Quienes practican ese juego olvidan que un libro es más que una estructura verbal, o que una serie de estructuras verbales; es el diálogo que entabla con su lector y la entonación que impone a su voz y las cambiantes y durables imágenes que deja en su memoria[3]

Borges is criticizing in the context of the Avant-garde a certain degree of arbitrary which was also responsible for the end of Avant-garde literature and art. Plus, he is pointing out, that the process of reading is not countable or measurable in all its parts, because it is a dynamic dialog between each reader and the text. Nevertheless, the idea of playing a game of combinatorics when constructing a text is quite common in the context of Avant-garde literature as well as in the context of emblematic art/poetry and mannerism. If it is a game, it should have rules that can be described.

From a mathematical point of view, combinatorics do not model rational sequences but countable structures. Applying this model on the literary text, it seems possible to detect the rules behind the structure. If we define a text as a combination of elements (letters, words, lemma, interpunction, POS, n-grams, etc.) we can count these elements and compare texts to find a pattern which might be characteristic for authorial style or genre. That is in a very reduced way an explanation of what stylometry is about. What we are not looking at, is the sense of the text. The semantic part of it is explicitly taken out of our view. That point is very important when talking about authors known for their pleasure in obscuring sense. Probably on another level, in the structure of combining elements, there is a pattern telling us some-

[2] See also on this subject: Christoph Strosetzki, "Von der Metaphysik-und Sprachkritik des Wiener Kreises zu Jorge Luis Borges und Ernesto Sábato", in *Wort und Zahl = Palabra y número* (Heidelberg, 2015), 289–300 and Volker Roloff, "Streifzüge durch surreale Bibliotheken: von Rabelais und Borges zum Internet", in *Durchquerungen: Ralf Schnell zum 65. Geburtstag*, eds. Iris Hermann und Anne-Maximiliane Jäger-Gogoll (Heidelberg: Winter, 2008) 225–34.

[3] Jorge Luis Borges, "Nota sobre (hacia) Bernhard Shaw", in *Otras inquisiciones* (1952), 93, https://apuntesliterarios.files.wordpress.com/2013/09/borges_otras_inquisiciones.pdf.

thing about the way these authors are constructing their texts ingeniously. The appalling effect of the "unreadable" text is already known but we would like to detect a rule (if there is one), like a new grammar, a solution of the combinatoric game. One could reasonably object that authors like Góngora and Picasso are probably pushing language intentionally to its limits where no rule or grammar is detectable, so why looking for it. But even if we take this as a hypothesis it would be nice to prove it with more evidence than our pure reading. Second hypothesis would be, that Góngora is rather using a kind of grammar or rule (we know already that he uses hyperbaton etc. to create new effects of language and grammar)[4] than Picasso. That is no revolting news to be proved but in comparison to other authors of their time it tells us about the influences and depending on the method also about how the poetry is constructed (apart from semantic riddles).

Nevertheless, it is obvious that no analysis can be free of the semantic level. When we look at the word which is used most in a corpus, we are reflecting about the "what does that mean?" question. And even before that: when we select our texts for the corpus decisions are made, based on the context knowledge, based on the semantics of each text. There is no claim of objectivity by using digital methods but it is in fact a completely different kind of looking at our texts and usually the path of knowledge becomes traceable.

Without the intention of presenting a solution, but rather with the intention of presenting various possibilities of stylometric analysis and discussing the advantages and disadvantages, I tested three different methods.

1. For microanalysis there would be POS (Part-of-speech-tagging) which tells us the concrete grammatical order of each sentence and how many sentences are structured the same way, etc.
2. For macroanalysis and a comparative view: Stylometry. Here it does not make sense to compare Góngora and Picasso directly because the corpora should be homogeneous regarding the period, number of words and genre.[5] But, we will see if it is possible to create two corpora for each author in its time and then to compare the results.

[4] For a detailed analysis of Góngora inventing a new language see Mercedes Blanco, *Góngora o La invención de una lengua, Lectura y signo* (León: Universidad de León, 2012).

[5] Cf. Christof Schöch, "Corneille, Molière et les autres: stilometrische Analysen zu Autorschaft und Gattungszugehörigkeit im französischen Theater der Klassik", *PhiN* 7 (2014): 130–57, http://web.fu-berlin.de/phin/beiheft7/b7t08.pdf; Maciej Eder, "Does size matter? Authorship attribution, short samples, big problem", in *Digital Humanities 2010: Conference Abstracts*, 2010, 132–5.

3. For a look into the sematic level, but without the usual given preferences, one could try Topic Modeling or Word Embeddings (word2vec). Here only Topic Modeling is tested rudimentarily for Picasso.

Before comparing and testing these three digital methods we should have a look on how traditional stylistics would proceed. In a rather one-sided, reductive way Sowinski is looking on Spitzer and stylistics from an actual point of view. His most important reproach is the intuitive way of analyzing literature:

> 1. intuitive Detailbetrachtung;
> 2. Feststellung von Gemeinsamkeiten im scheinbar Zufälligem;
> 3. Rückschluss auf den Seelenzustand des Verfassers beim nochmaligen Lesen des Ganzen[6]

Reading means here: intuitive micro-analysis and observation of details, which catches our individual interest and next step would be to connect some parts of the text based on arbitrary found similarities. In a last step all this is transferred to the complete work and soul of the author in describing his personal style. We know that Spitzer did not analyze texts in this simple way, and that stylistics today are used to distinguish our 'digital' selves from an intuitive way of reading. Though this image of stylistics might be the reason why it is no longer the state of the art in literary criticism.

Nevertheless, there are quite similar reproaches against stylistic analysis today:

> [stylistics] still operates according to the principle prove by examples. It should by now be clear that under such a regimen any theory can be proven right: one can always find an example that illustrates a theory, however outlandish it maybe.[7]

The solution for van Peer, Zyngier and Hakemulder seems to be statistical tests which are quite easy to elaborate for stylistics. But, proving theories with statistical values does not automatically mean that they are more valid than those who are using other methods and models. It always depends on the choice you make what to count and measure. It is true though that the

[6] Bernhard Sowinski, *Stilistik: Stiltheorien und Stilanalysen*, 2nd rev. ed., Sammlung Metzler 263 (Stuttgart: Metzler, 1999), 138.

[7] Willie van Peer, Sonia Zyngier, and Jèmeljan Hakemulder, "Foregrounding, Estrangement and Pattern", in *Stylistics: Prospect & Retrospect*, ed. by David L. Hoover, PALA Papers 3 (Amsterdam: Rodopi, 2007), 1–22, here 5.

chance to test arguments with statistics makes your hypotheses verifiable and if providing the used data also intersubjectively verifiable.

Still, we should not forget that in literary criticism one could never measure statistically all factors of theory. For example, for the concept of intertextuality each text is the result of many others plus the predisposition of the reader. For Borges the identity of the author, his soul or his style is not at all important because:

> El hecho es que cada escritor crea sus precursores. Su labor modifica nuestra concepción del pasado, como ha de modificar el futuro. En esta correalación nada importa la identidad o la pluralidad de los hombres.[8]

## Measuring and Tagging Picasso

Both, Picasso and Góngora are using prepositions (most frequently "de") to construct sequences that seem to be readable backwards as well as in the given order (forwards). A very often cited example for Góngora is "montes de agua y piélagos de montes"[9] which describes the disordered perception of the young shipwrecked who was just washed up the beach at the beginning of *Soledades*. We are not sure if he describes the giant waves as mountains or the giant mountains as waves or sea in general.[10] How could we ever represent (or model) an effect of reader-irritation and uncertainty of sense in a quantitative and measurable way? Probably not the effect itself, but the significant frequency of the same grammatical construction could show us the pattern of a strategy used by the author all over his work. While, the examination and interpretation of several examples explains the effect itself and its meaning.[11]

For Picasso we do not have very well analyzed examples. Therefore, we draft a possible way of how methods like POS-tagging could work but with-

[8] Jorge Luis Borges, "Kafka y sus precursores", in *Obras completas* (Buenos Aires 1974), 711–12.

[9] Luis de Góngora, *Soledades*, ed. by John Beverley (Mardid: Catedra, [14]2009), V. 44, 77.

[10] See for more details Bernhard Teuber, "Curiositas et crudelitas: das Unheimliche am Barock bei Góngora, Sor

Juana Inés de la Cruz und José Lezama Lima", in *Diskurse des Barock: dezentrierte oder rezentrierte Welt?*, ed. by Joachim Küpper and Friedrich Wolfzettel (München: Fink, 2000), 615–52 and Rißler-Pipka, *Picassos schriftstellerisches Werk*, 261–91.

[11] Both parts of analysis, the quantitative and the hermeneutic qualitative are shown in the study by Marie-Eglantine Lescasse, "Góngora hors norme? Étude stylométrique d'un motif góngorin", in this volume, p. 91.

out really proceeding them seriously to come to results.[12] A starting point for the argumentation is the hypothesis that Picasso's writings are extremely unconventional even for his period, the avant-garde. That is one of the common ways of arguing in the context of literary history that the author X has an outstanding individual style which needs more detailed analysis than others and is worth a redefinition of the canon.

As we already know, that Picasso does not write in sentences (he is rejecting punctuation and other grammatical order). There is the possibility that he creates his own grammatical rule or his own language. To detect the order on the level of microanalysis Part-of-speech-tagging is a common method in linguistics. What we would like to count is the number of prepositions, verbs, nouns, adverbs, personal pronouns, etc. to compare this number with other authors. Picasso's system of connecting the disparate in a baroque manner is completely dependent on prepositions and pronouns:[13]

> la ilusión **del** drama **que se** juega la vida **sobre** la última carta
> y hace **su** castillo **de** naipes **sobre** el filo **de** la navaja **del** son
> **del** canto **de** cisne **del** clarín **que se** muere **de** pena nadando
> buzo ciego **por** las tripas verdes **del** lago **por donde** Ofelia va
> buscando **en** el fondo un trozo **de** papel y un lápiz

Reading this rather randomly chosen verses of a longer prose poem (dated 02.02.1936), Picasso's style is intuitively detectable as a form of adding new aspects to a poetic image by enlarging a never-ending sentence. We understand motives like the Ophelia myth (the lake, the swan, swimming, diving) or theatre (illusion, drama, playing, life, song, horn ("clarín"), probably also the house of cards and dying of pain) or writing/painting (letter, poetry ("canto" can mean song and poetry as well), the colour "green", sheet of paper and pencil). The remaining nouns which are more difficult to understand are knife edge ("el filo de la navaja") and tripe ("tripas"). They are part of the kitchen and food vocabulary which has an important role in Picasso's writings and at once belong to the leitmotif of bullfight.[14] Analyzing his motives and trace them back to predecessors like Góngora can be done by examples

[12] That would be another study to do. This paper is a reflection about the possible methods of stylometry in comparison to known theories of authorship and stylistics.

[13] Pablo Picasso, *Écrits*, ed. by Christine Piot and Marie-Laure Bernadac (Paris: Gallimard, 1989), 100–1 [Highlighting of prepositions and pronouns by N. R-P.].

[14] Particularly the tripe ("tripas") of the horse in the bullfight which is visible when the bull cuts the stomach of the horse with his horn is a very known and discussed motive in Pi-

like this citation, because the very significant point about Picasso's writing is, that he is repeating these motives and vocabulary over and over again in 200 prose poems (60.805 words) written in Spanish and 346 prose poems (56.518 words) written in French.[15]

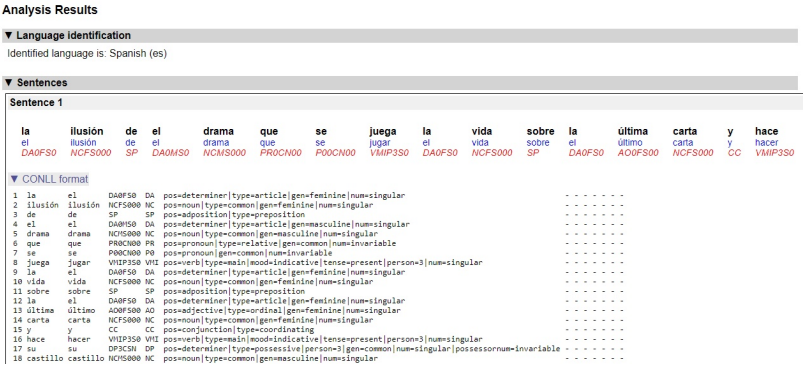
Though, what we cannot detect by this kind of hermeneutic, stylistic (qualitative) analysis is the function of the quantitative significant number of prepositions and pronouns. As we see in the few verses, the prepositions and pronouns are responsible for the loss of sense. There is no semantic link between the motives like "castillo de naipes" and "el filo de la navaja" and "son del canto" and "cisne" and "clarín" – if we cut the prepositions and pronouns out ("su, sobre, de, del") and leave only those who are part of a collocation. There is still a link between "son del canto" and "clarín" but because of the insertion of "de cisne del" it becomes: the sound of the song of the swan of the horn – and this is obviously a very broken and crooked image. We are literary forced to try to make sense out of the bizarre combination of motives Picasso is presenting us because he collates them via prepositions and pronouns – a technique very similar to collage and montage in plastic arts as well as in avant-garde literature. All this should be reason enough for Part of Speech Tagging. To test it, without much technical effort and knowledge in linguistic tools, I tried the demo version of FreeLing[16] with the same verses (cited above).

For the few lines of verse (67 words = token) we get as an output a list of part of speech-tags for each token and an even more detailed XML-file (11 pages) where each token gets a description, see fig. 1.

casso's painting and writing. See Rißler-Pipka, *Picassos schriftstellerisches Werk*; Marie-Laure Bernadac, "Le Gazpcho de la Corrida", in *Picasso, toros y toreros* (Paris: RMN, Seuil, 1993), 47–59.

[15] For an overview and comparison see Nanette Rißler-Pipka, "Picasso et son esthétique numérique", in *Zwischen Genres und Medien: Formen moderner Prosa in Frankreich*, ed. by Christof Schöch and Nanette Rißler-Pipka, *PhiN Beiheft* 16 (2019), http://web.fu-berlin.de/phin/beiheft16/b16t04.pdf.

[16] FreeLing is created in Spain and therefore probably more adopted to Spanish language than other POS tools. For further information about the tool: Lluís Padró and Evgeny Stanilovsky, "FreeLing 3.0: Towards Wider Multilinguality", in *Proceedings of the Language Resources and Evaluation Conference (LREC 2012)* (Istanbul, Turkey: ELRA, 2012), http://nlp.lsi.upc.edu/publications/papers/padro12.pdf.



**Analysis Results**

▼ **Language identification**

Identified language is: Spanish (es)

▼ **Sentences**

**Sentence 1**

| la | ilusión | de | el | drama | que | se | juega | la | vida | sobre | la | última | carta | y | hace |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| el | ilusión | de | el | drama | que | se | jugar | el | vida | sobre | el | último | carta | y | hacer |
| *DA0FS0* | *NCFS000* | *SP* | *DA0MS0* | *NCMS000* | *PR0CN00* | *P00CN00* | *VMIP3S0* | *DA0FS0* | *NCFS000* | *SP* | *DA0FS0* | *AO0FS00* | *NCFS000* | *CC* | *VMIP3S0* |

▼ CONLL format

```
1  la        el        DA0FS0  DA  pos=determiner|type=article|gen=feminine|num=singular                    - - - - - - -
2  ilusión   ilusión   NCFS000 NC  pos=noun|type=common|gen=feminine|num=singular                          - - - - - - -
3  de        de        SP      SP  pos=adposition|type=preposition                                         - - - - - - -
4  el        el        DA0MS0  DA  pos=determiner|type=article|gen=masculine|num=singular                  - - - - - - -
5  drama     drama     NCMS000 NC  pos=noun|type=common|gen=masculine|num=singular                        - - - - - - -
6  que       que       PR0CN00 PR  pos=pronoun|type=relative|gen=common|num=invariable                    - - - - - - -
7  se        se        P00CN00 P0  pos=pronoun|gen=common|num=invariable                                  - - - - - - -
8  juega     jugar     VMIP3S0 VMI pos=verb|type=main|mood=indicative|tense=present|person=3|num=singular - - - - - - -
9  la        el        DA0FS0  DA  pos=determiner|type=article|gen=feminine|num=singular                  - - - - - - -
10 vida      vida      NCFS000 NC  pos=noun|type=common|gen=feminine|num=singular                        - - - - - - -
11 sobre     sobre     SP      SP  pos=adposition|type=preposition                                         - - - - - - -
12 la        el        DA0FS0  DA  pos=determiner|type=article|gen=feminine|num=singular                  - - - - - - -
13 última    último    AO0FS00 AO  pos=adjective|type=ordinal|gen=feminine|num=singular                   - - - - - - -
14 carta     carta     NCFS000 NC  pos=noun|type=common|gen=feminine|num=singular                        - - - - - - -
15 y         y         CC      CC  pos=conjunction|type=coordinating                                       - - - - - - -
16 hace      hacer     VMIP3S0 VMI pos=verb|type=main|mood=indicative|tense=present|person=3|num=singular - - - - - - -
17 su        su        DP3CSN  DP  pos=determiner|type=possessive|person=3|gen=common|num=singular|possessornum=invariable - - - - - - -
18 castillo  castillo  NCMS000 NC  pos=noun|type=common|gen=masculine|num=singular                        - - - - - - -
```

Figure 1: Screenshot of FreeLing-Results (http://nlp.lsi.upc.edu/freeling/demo/demo.php) and a transcription of line 1-5

```
1  la        el        DA0FS0  DA  pos=determiner|type=article|gen=feminine|num=singular  - - - - - - -
2  ilusión   ilusión   NCFS000 NC  pos=noun|type=common|gen=feminine|num=singular          - - - - - - -
3  de        de        SP      SP  pos=adposition|type=preposition                         - - - - - - -
4  el        el        DA0MS0  DA  pos=determiner|type=article|gen=masculine|num=singular  - - - - - - -
5  drama     drama     NCMS000 NC  pos=noun|type=common|gen=masculine|num=singular        - - - - - - -
```

```
<document>
  <wordcount>67</wordcount>
  <cputime>0.017303</cputime>
  <paragraph>
    <sentence id="1">
      <token begin="0" ctag="DA" end="2" form="la" gen="feminine" id="t1.1"
lemma="el" num="singular" pos="determiner" tag="DA0FS0" type="article">
        <morpho>
          <analysis ctag="DA" gen="feminine" lemma="el" num="singular"
pos="determiner" selected="1" tag="DA0FS0" type="article"/>
          <analysis case="accusative" ctag="PP" gen="feminine" lemma="lo"
num="singular" person="3" pos="pronoun" tag="PP3FSA0" type="personal"/>
          <analysis ctag="NC" gen="masculine" lemma="la" num="singular"
pos="noun" tag="NCMS000" type="common"/>
        </morpho>
      </token>
```

Table 1: xml-document, result of POS-Tagging via FreeLing (detail for the first word “la”)

All this is to describe the document, the sentence (if we can call it one) and the first tiny word “la”. We now have metadata about the word “la” and in whole about the few verses of Picasso we analyzed by using POS-Tagging. Let’s assume, that we would have all poetry by Picasso in XML-format and already tagged, we could look out for the pattern of grammatical construction – even if each poem (some of them consist of more than 6000 words) is one long sentence. Questions could be answered like: How long are the sequences of noun preposition noun (probably added a DA (determiner, article)? Yet, by looking a bit deeper in the analysis of our few verses, we can already make some significant observations.

Sequences look here like: DA NC SP DA NC PR PO (see the example above). That means there are many long sequences without a single verb. Out of the 67 words in the cited verses, we find only 4 conjugated verbs (juega, hace, muere, va – all 3. person sing.) and 2 in gerund forms (nadando, buscando). At least this is what we understand as a human being reading the text. The machine shows us other possibilities of understanding the pure grammatical form and produces without intention an alienation effect.[17] Of course, we know that words can represent different grammatical types, but are usually clearly marked by their context. For example, the difference between "el juego" (the game) and "juego al fútbol" (I play soccer) is easily detectable by the article "el".

Through the process of POS-Tagging, we get each word analyzed separately and shown all possible grammatical meanings. You can see it above in the simple example of "la": tagged correctly as DA0FS0: determiner, article, feminine, singular. But, in the following rows it shows also the other possible meanings: PP3FSA0: personal pronoun, third person, feminine, singular, accusative, or another possible meaning: NCMS000: noun, common, masculine, singular.[18] All these alternative meanings are excluded for our human reader-eye because we cannot ignore context and its meaning and probably do not know all possible grammatical meanings at once. For Picasso and the poetics of Avant-garde, it would be real fun to be able to imagine every possible meaning of every single word at once.

The fact, that Picasso as an author of avant-garde poetry is getting nearly that far can be proved by POS-Tagging because it is his authorial style of constructing sequences with prepositions that produces failures in the functionality of the tool. I rather would not have recognized it because reading all the 11 pages of tags in the XML-document is not that amusing. But, when I was looking at the possible forms of verbs (because they are obviously rare as we see by reading the 67 words: 4 verbs, 2 gerund forms), I was surprised by the tag for "son": VSIP3P0: verb, semiauxiliary, indicative, present, 3. person, plural.

[17] This is exactly the same effect, Stephen Ramsay is speaking about in, *Reading machines: toward an algorithmic criticism, Topics in the digital humanities* (Urbana: University of Illinois Press, 2011).

[18] For all, who are like myself not Spanish native speaker: "la" has also the meaning of the "A", a musical note. See "La = Cf. fa. 1. m. Sexta nota de la escala musical." (*Diccionario de la lengua española*, Real Academia Española, http://dle.rae.es/?id=MiXNIfB|MiXQLcB|MiZ5vEt).

```
</token>
      <token begin="124" ctag="VSI" end="127" form="son" id="t1.29"
lemma="ser" mood="indicative" num="plural" person="3" pos="verb" tag="VSIP3P0"
tense="present" type="semiauxiliary">
        <morpho>
          <analysis ctag="VSI" lemma="ser" mood="indicative" num="plural"
person="3" pos="verb" selected="1" tag="VSIP3P0" tense="present"
type="semiauxiliary"/>
          <analysis ctag="NC" gen="masculine" lemma="son" num="singular"
pos="noun" tag="NCMS000" type="common"/>
        </morpho>
      </token>
```

Table 2: xml-document, result of POS-Tagging via FreeLing (detail for the 27th word "son") and transcription of the tag "VSIP3PO VSI"

The context of the word "son" makes it very clear, that Picasso is talking about sound: "el son del canto de cisne del clarín". The result of POS-Tagging is, that "son" means: 3. person plural of the verb "ser", which is obviously wrong – even if we assume that Picasso is linking words in an alienated way. So, why did the POS-Tagging identify the word wrongly? Probably because Picasso's grammatical construction, the sequence of noun, preposition, noun, etc. is not the usual way how language is used.[19]

Three consequences for a possible POS-Tagging-analyses of Picasso's writing should be noted: First, the reliance of the tool cannot be guaranteed. Nobody would recognize failures like this one in about 60.000 words. Though, for an overview of the quantity of long sequences of preposition-noun-etc. the tool will certainly give reliable results. Second, the unexpected and very powerful result is the detection of these failures as evidence for Picasso's outstanding style. Naturally, this could also be the effect of a failure in the programming of the tool which I could not test but other examples show it is not an exception (see Tab. 3). Third, the listed possibility of meaning due to the grammatical function of each word show us the uncertainty and vagueness of sense in Picasso's writings. For this last point we also find an example in our few verses:

[19] The ambiguity of POS-Tagging in poetry is also verifiable in Spanish Golden Age sonnets, see Borja Navarro-Colorado, "A Metrical Scansion System for Fixed-Metre Spanish Poetry", *Digital Scholarship in the Humanities* 33, no. 1 (1. April 2018): 112–27, https://doi.org/10.1093/llc/fqx009.

```
</token>
      <token begin="187" ctag="NC" end="191" form="buzo" gen="masculine"
id="t1.44" lemma="buzo" num="singular" pos="noun" tag="NCMS000" type="common">
        <morpho>
          <analysis ctag="NC" gen="masculine" lemma="buzo" num="singular"
pos="noun" selected="1" tag="NCMS000" type="common"/>
          <analysis ctag="VMI" lemma="buzar" mood="indicative" num="singular"
person="1" pos="verb" tag="VMIP1S0" tense="present" type="main"/>
        </morpho>
      </token>
      <token begin="192" ctag="AQ" end="197" form="ciego" gen="masculine"
id="t1.45" lemma="ciego" num="singular" pos="adjective" tag="AQ0MS00"
type="qualificative">
        <morpho>
          <analysis ctag="AQ" gen="masculine" lemma="ciego" num="singular"
pos="adjective" selected="1" tag="AQ0MS00" type="qualificative"/>
          <analysis ctag="VMI" lemma="cegar" mood="indicative" num="singular"
person="1" pos="verb" tag="VMIP1S0" tense="present" type="main"/>
          <analysis ctag="NC" gen="masculine" lemma="ciego" num="singular"
pos="noun" tag="NCMS000" type="common"/>
        </morpho>
      </token>
```

Table 3: xml-document, result of POS-Tagging via FreeLing (detail for the 44th and 45th words "buzo" and "ciego") and transcription of the tags "NCMS000 NC", "AQ0MS00 AQ"

Reading the part of the cited verses: "nadando buzo ciego por las tripas…" we cannot be sure if a blind diver is swimming to find the green tripe or if some kind of narrator (yo buzo) is diving blind and at once swimming to find the green tripe. The word "buzo" can mean the first person singular of the verb "buzar" as well as the noun "el buzo". In the context of the poem and the construction of the sentence (or sequence) we are simply not able to decide which one of the meaning should be favored. The POS-tagging gives us all possible meaning of each words – at least regarding its grammatical function. That Picasso is intentionally playing with the grammatical scope and indeterminateness is now not only an intuitive hypothesis while reading his text but can easily be argued with the help of POS-tagging.[20]

## Stylometric experiments for Góngora (and Picasso)

Stylometry is an important and increasing field in quantitative literary analysis.[21] It is one of the main subjects of this volume (see the contributions of Fradejas Rueda, Hernández Lorenzo, Calvo Tello, Lescasse). For the Spanish speaking world José Calvo Tello explained the functionality of Delta and

[20] From a linguistic point of view, the semantics of poetry are rarely analyzed and can also be a connection between close reading and digital methods, see Aurélie Herbelot, "The Semantics of Poetry: A Distributional Reading", *Digital Scholarship in the Humanities* 30, no. 4 (Dezember 2015): 516–31, https://doi.org/10.1093/llc/fqu035.

[21] See the Special Interest Group (SIG-DLS) for Digital Literary Stylistics: https://dls.hypotheses.org/ and the new priority programm for Computational Literary Studies of the DFG: http://www.dfg.de/en/service/press/press_releases/2018/press_release_no_07/index.html

Christof Schöch explained in the introduction into *Digital Humanities* quantitative literary analysis in its different methods.[22] Historically born out of authorship contribution and forensic linguistics,[23] stylometry is now used for different kind of literary research questions (re-writing literary history, changes of style within the works of one author, multilingual comparisons, genre classification, collaborative authorship). Instead of supporting the positivistic thesis that every author has an individual style, that can be proved like a fingerprint, more and more analysis in the literary context show that stylistic differences can be due to genre signals and others.

To compare Góngora and Picasso each of them to contemporary authors of approximately the same genre I used *stylo*[24], a package for the statistical program *R*, that calculates the statistical value for the 'stylistic' distance for each text of the corpus to another (based on the frequencies in the whole corpus). That means, the result depends enormously on the corpus composition, on the chosen parameters (how many frequent words are taken into consideration, etc.) and on the chosen delta (Euclidian, cosine, Eder's, Burrow's, etc.).[25] The advantage of such a tool is the very transparent decision (all parameters are given in the visualization of the results) and that it is rather easy to understand – even if the mathematics behind it are quite complicated.

> While conceptually straightforward, stylometry is rather advanced when it comes to its mathematical background; simultaneously comparing several frequencies of function words requires techniques which are referred to as

[22] José Calvo Taller, "Entiendiendo Delta desde las Humanidades", *Caracteres, Estudios culturales y críticos de la esfera digital*, 5(1): 140–76, http://revistacaracteres.net/revista/vol5n1mayo2016/entendiendo-delta/; Christof Schöch, "Quantitative Analyse", in *Digital Humanities*, ed. by Fotis Jannidis, Hubertus Kohle and Malte Rehbein (Stuttgart: J.B. Metzler, 2017), 279–98, https://doi.org/10.1007/978-3-476-05446-3_20.

[23] For early spanish literature see Javier Blasco and Cristina Ruíz Urbón, "Evaluación y cuantificación de algunas técnicas de 'Atribución de autoría' en textos españoles", *Castilla. Estudios de Literatura* (15. Oktober 2009): 27–47, https://doi.org/10.24197/cel.0.2009.27-47; Javier Blasco Pascual, Patricia Marín Cepeda and Cristina Ruiz Urbón, Ed., *Hos ego versiculos feci: estudios de atribución y plagio* (Madrid and Frankfurt am Main: Iberoamericana and Vervuert, 2010); Nanette Rißler-Pipka, "Digital Humanities und die Romanische Literaturwissenschaft: der Autorschaftsstreit um den *Lazarillo de Tormes*", *Romanische Forschungen* 128, no. 3 (15. September 2016): 316–42.

[24] Maciej Eder, Mike Kestemont and Jan Rybicki, "Stylometry with R: a suite of tools", in *Digital Humanities 2013*. Conference Abstracts (Lincoln: University of Nebraska-Lincoln, 2013), 487–89.

[25] For a list of all parameters, see Maciej Eder, Jan Rybicki and Mike Kestemont, "Stylo: A Package for Stylometric Analyses", last modified June 2017, https://goo.gl/oZ5bBh.

> multivariate, or multidimensional, because they involve multidimensional geometry to compute similarities between texts.[26]

In contrast to traditional stylistics, the choice of what should be the basis of comparison is not made by the researcher (for example the comparison of a very typical word or sequence) but is simply based on the MFW (Most Frequent Words). Of course, by changing the parameters and picking only the significant results, one can still manipulate the outcome. Yet, the changing of parameters has also the character of experimenting and if every decision is transparent and discussed it can be part of a positive research process.

In the era of digitalization, it is still not at all natural to find a digitized corpus with reliable texts and metadata. For the 17th century Borja Navarro provides a valuable resource of Spanish baroque sonnets, including Góngora.[27] The problem is, each sonnet is given as a single xml-file. While, for stylometry I need more words and txt-format (xml is also possible, but the real problem are the small single files with limited number of token (words)). That is why in the end I took the texts directly from the Biblioteca Virtual Miguel de Cervantes (BVMC) to fold the sonnets for each author into one single document, so we get files of 3000-5000 words per author. This is about the minimum we need to proceed with stylometry with R. What really could be a minimum of words used in *stylo* is discussed by Maciej Eder. He also points out, that short literary forms like sonnets or aphorisms in his example have a very genre driven style and it is problematic to treat many short texts of this kind as one long one.

> Short literary forms are often masterpieces of concise language, with a domination of verbs over adjectives, particles and so on, with a proverbial witty

[26] James O'Sullivan, Katarzyna Bazarnik, Maciej Eder and Jan Rybicki, "Measuring Joycean Influences on Flann O'Brien", *Digital Studies = Le Champ Numérique* 8, no. 1 (27. März 2018), https://doi.org/10.16995/dscn.288.

[27] Borja Navarro, "A computational linguistic approach to Spanish Golden Age Sonnets: metrical and semantic aspects", in *Proceedings of the Fourth Workshop on Computational Linguistics for Literature*, 2015, 105–113, http://aclweb.org/anthology/W15-0712 and for the corpus itself: *Corpus of Spanish Golden-Age Sonnets*, ed. by Borja Navarro Colorado, María Ribes Lafoz and Noelia Sánchez, 2015. https://github.com/bncolorado/CorpusSonetosSigloDeOro : see also the review: José Calvo Tello, "Corpus of Spanish Golden-Age Sonnets", in *RIDE: a review journal for digital editions and resources* 6 (September 2017), https://ride.i-d-e.de/issues/issue-6/corpus-of-spanish-golden-age-sonnets/. The recent project Diachronic Spanish Sonnet Corpus (DISCO) enlarges this work by providing more than 3000 sonnets from 15th–19th century: Ruiz Fabo, Pablo, Helena Bermúdez Sabel, Clara Martínez Cantón and José Calvo Tello, *Diachronic Spanish Sonnet Corpus (DISCO)*, (Madrid: UNED, 2017). https://github.com/pruizf/disco.

> style, and with a strong tendency to compression of content. Thus, a collection of aphorisms will certainly have different word frequencies than a long essay written by the same author[28]

The next problem is, that now we have one single 'text' per author and the typical dendrogram as a statistical output of *stylo* tries to cluster at least pairs of texts. Usually, it is better to have at least two texts by the same author to allow a correct clustering. So, if we have 7 authors (that was the number of authors with enough text, easily accessible via BVMC, and from the baroque period) one must be the outlier. And it is not surprisingly Góngora in the first experiment (fig.2[29]).

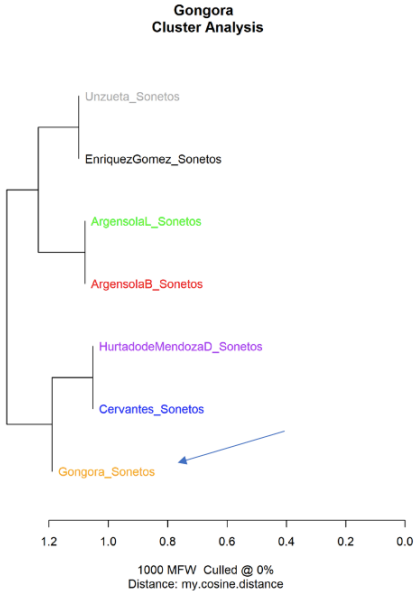



Figure 2: Dendrogram of the Corpus of Spanish Golden Age Sonnets (1000 MFW, cosine Delta), arrow added to point out Góngora

[28] Maciej Eder, "Does size matter? Authorship attribution, short samples, big problem".

[29] The result didn't change when changing the parameters from 100-1000 MFW or using Eder's distance, but as the cosine distance was tested to perform best with literary texts in various languages, I used only cosine in the following experiments, see: Fotis Jannidis et al., "Improving Burrows' Delta: an empirical evaluation of text distance measures", in *Digital Humanities 2015*: Conference Abstracts (Sydney, 2015), http://dh2015.org/abstracts/index.php.

Looking out for more texts to enlarge the corpus, I added 25 authors to have an even number of 32 in total to pair the authors in the dendrogram. But, still there is only one text per author (see fig. 3). The choice of the additional authors was made by the help of Borja Navarro's collection of Golden Age sonnets and the precondition of finding enough sonnets (at least 40) in BVMC. Furthermore, the difference between the number of words for each text (collection of sonnets) varies between 1850 (Tassis y Peralta) and 13061 (Lope de Vega). Another problem is the wide range of age: Some of the authors are more than hundred years younger/older than others (therefore I highlighted the dates of authors who died in 16th century in the table 4).

Table 4: Table of used corpus

| Author | Number of words | Biographical data | File-name |
|---|---|---|---|
| Hernando de Acuña | 4093 | 1518–1580 | AcunyaF_Sonetos |
| Francisco de Aldana | 4497 | 1537–1578 | AldanaF_Sonetos |
| Juan de Almeida | 4292 | 1530–1573 | AlmeidaJ_Sonetos |
| Bartolomé Leonardo de Argensola | 3334 | 1562–1631 | ArgensolaB_Sonetos |
| Lupercio Leonardo de Argensola | 2830 | 1559–1613 | ArgensolaL_Sonetos |
| Francisco de Borja y Aragón | 2935 | 1577–1658 | BorjayAragonF_Sonetos |
| Juan Boscán Almogáver | 4975 | 1492–1542 | BoscánAlmogáverJ_Sonetos |
| Luis Carrillo y Sotomayor | 2551 | 1582–1610 | CarrilloySotomayorL_Sonetos |
| Miguel de Cervantes Saavedra | 3604 | 1547–1616 | Cervantes_Sonetos |
| Gutierre de Cetina | 3262 | 1510–1554 | CetinaG_Sonetos |
| Antonio Enriquez Gómez | 4216 | 1600–1661 | EnriquezGomez_Sonetos |
| Francisco de Trillo y Figueroa | 4871 | 1618–1680 | FigueroaF_Sonetos |
| Garcilaso de la Vega | 3787 | 1501–1536 | GarcilasodelaVega_Sonetos |
| Luis de Góngora y Argote | 2936 | 1561–1627 | Gongora_Sonetos |
| Fernando de Herrera | 6165 | 1534–1597 | HerreraF_Sonetos |
| Diego Hurtado de Mendoza | 5361 | 1503–1575 | HurtadodeMendozaD_Sonetos |
| Fray Luis de Léon | 6330 | 1527–1591 | LeonFrayLuis_Sonetos+Eglogas |
| Joseph de Litala y Castelvi | 4891 | 17th century | LitalayCastelviJ_Sonetos |
| Lope de Vega | 13061 | 1562–1632 | LopedeVega_Sonetos |
| Francisco de Medrano | 5188 | 1570–1607 | MedranoF_Sonetos |

| Author | Number of words | Biographical data | File-name |
|---|---|---|---|
| Pedro de Padilla | 4245 | 1540–1595 | PadillaP_Sonetos |
| Bernardino de Rebolledo | 2418 | 1597–1674 | RebolledoB_Sonetos |
| Agustín de Salazar y Torres | 3342 | 1642–1675 | SalazarYTorresA_Sonetos |
| Sor Juana Inés de la Cruz | 3430 | 1648–1695 | SorJuanaInesdelaCruz_Sonetos |
| Pedro Soto de Rojas | 3635 | 1584–1658 | SotodeRojasP_Sonetos |
| Juan de Tassis y Peralta | 1850 | 1582–1622 | TassisyPeraltaJ_Sonetos |
| Juan Timoneda | 2688 | 1520–1583 | TimonedaJ_Sonetos |
| Tirso de Molina | 5615 | 1579–1648 | TirsodeMolina_Sonetos |
| Francisco de la Torre | 3723 | 1521–1582 | TorreF_Sonetos |
| Luis de Ulloa Pereira | 3478 | 1584–1674 | UlloaPereiraL_Sonetos |
| Gabriel Bocángel y Unzueta | 3965 | 1603–1658 | Unzueta_Sonetos |
| Diego Ximénez Ayllón | 2660 | 1530–1590 | XimenezAyllon_Sonetos |

In the comparison of 32 authors Góngora is not at all an outlier (at least not in the genre of sonnets). He is prominently clustered with Lope de Vega, and on a sidetrack with the rather unknown Joseph (or José) de Litala y Castelvi (a knight in the Order of Calatrava).[30] The clustering of Lope and Góngora is not that surprising even if they were literary enemies. Both wrote sonnets as satire to fight on the field of literature against each other. The different genres of sonnets (love, satire, religion, honor) are also a problem for a stylistic comparison via MFW.[31] More important than this aspect seems to be the distance of different periods: all authors on the lower arm of the dendrogram are 16th century authors. Only Francisco de Aldana and Fray Luis de Léon (clustered together), Francisco de la Torre and Fernando de Herrera (clustered together) are clustered on the upper arm of the dendrogram and died already at the end of 16th century. This phenomenon can be observed all along the following experiments (also in the series of 100–1000 MFW which

[30] In 1941 Homero Serís is trying to establish José Delitala y Castelvi as "un clásico olvidado", obviously without much success. Homero Serís, "Libro raro y curioso. Poesías de José Delitala y Castelví (1672). Un clásico olvidado", *Bulletin Hispanique* 43, no. 2 (1941): 171–81, https://doi.org/10.3406/hispa.1941.2908.

[31] As all sonnets are treated as one long text, the different genres are mixed together in one text. Therefore, also in the sonnet-collection of Góngora and Lope de Vega not only the satire sonnets are compared.

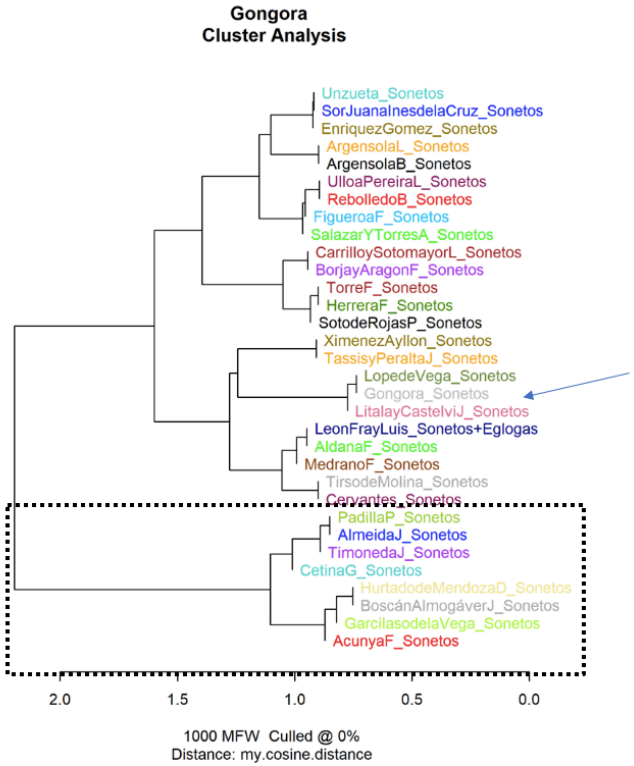



Figure 3: Dendrogram for an enlarged corpus for 32 authors of Golden Age sonnets, same parameters, arrow added to point out clustering of Góngora, Lope and Litala, colored box added to point out the arm of older sonnets

are not discussed here) and therefore is very consistent under different parameters.

For the next experiment, I changed the corpus another time: deleted Fray Luis de Léon, because he did not write enough sonnets (before I added his Eglogías to get the minimum number of words) and replaced him with Francisco de Quevedo, because he is also very well known for the satire sonnets in the literary battle between Góngora and other poets. Another important change is to split the sonnets for each author in at least two texts (of 1000–3000 words). The results now show the problems of the corpus and probably the difficulty of treating sonnets as one long text. In the series of 100–1000 MFW the clustering was not consistent and never could cluster all authors

correctly. The example of 1000 MFW here is picked out to have the same parameters as before (fig. 2–4).

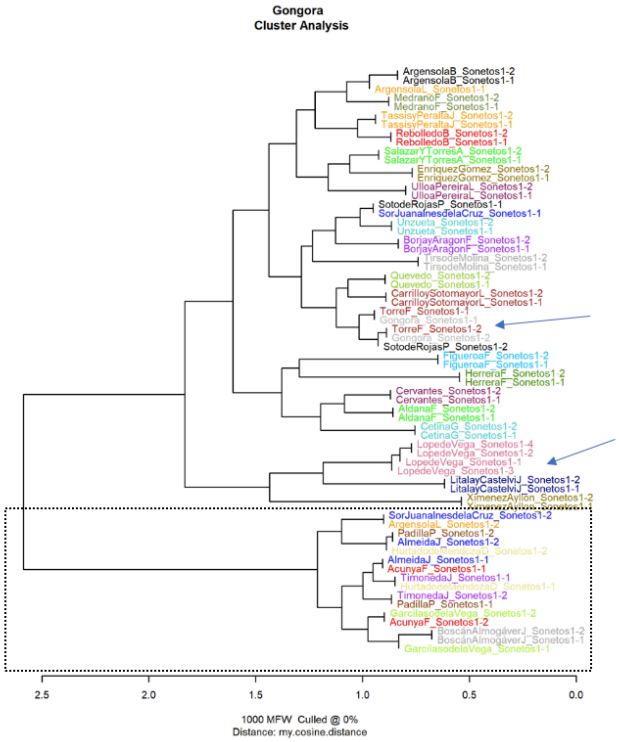



Figure 4: Dendrogram for sonnet-corpus split in at least two texts, same parameters; arrows added to point out clustering of Góngora, Lope and Litala, colored box added to point out the arm of older sonnets

Góngora is now split up and clustered with Francisco de la Torre and Pedro Soto de Rojas but sharing the arm of the dendrogram also with Quevedo (here similarities in style are known). In the whole picture, most of the authors are clustered correctly.[32] For example, Lope de Vega was split up in four texts (because of the enormous number of words) and all four found together on the same arm with the unknown Litala y Castelevi as before (fig. 2 and 4).

[32] "Correctly" means here: all parts of sonnets by the same author are clustered on one single arm of the dendrogram together.

Obviously wrong is the clustering of Sor Juana Inés de la Cruz and Lupercio Leonardo de Argensola, because one part of their sonnets was clustered in the period of 16th century authors and the other as before in the younger group. All in all, the 16th century arm of the dendrogram with Garcilaso de la Vega now is still detectable as a different style, but in itself only Juan Boscán Almogáver was clustered correctly.

This last experiment leads to two conclusions and further experiments: First, if the clustering of authorship basically works – even for a complex source like poetry – it might tell us something about stylistic difference of each author to another. Second, the inconsistencies in the corpus are probably due to the wide range of period and size (number of words). That means if we delete all texts of the 16th century and the ones less than 1000 words the results should be consistent. Before, we should have a look in the series of experiments *stylo* did already from 100 to 1000 MFW. If the number of words is part of the problem, with less MFW the performance is probably more consistent. But, as we see this is only true for the 17th century texts (fig. 5).

Now only 4 authors are not clustered correctly (in fig. 4 with 1000 MFW 10 authors are not clustered correctly) and Góngora is back on the arm of the dendrogram with Lope de Vega and Litala y Castelevi. This reappearance in different experiments gives more evidence for a stylistic similarity between the authors. But, it tells us nothing about the question if Góngora is the stylistic outlier who invented a new language. Also, in the following experiment, when we delete all sonnets from authors died before the 17th century and with less than 1000 words Góngora seems to be less an outlier than expected (fig. 6).

Now the outliers are Lope de Vega and Litala y Castelevi (both before clustered with Góngora). Apart of Sor Juana Inés de la Cruz and Lupercio Leonardo de Argensola all split sonnet texts are clustered correctly with their authors. Both, Sor Juana and L. Argensola were also in the first experiment with split sonnet-corpus wrongly clustered (fig. 4–5). This is probably due to the extremely mixed themes in their sonnets: Argensola's sonnets in the collection 1-1 are more about history and heroes and the collection 1-2 has religious subjects. For Sor Juana, we know that of course, many sonnets have religious subjects but are silently mixed with love-themes. In the collection of sonnets 1-2 of both authors (here clustered together) the word "pecho" (sin) is one of the most used content words. But, statistics reveal that it is

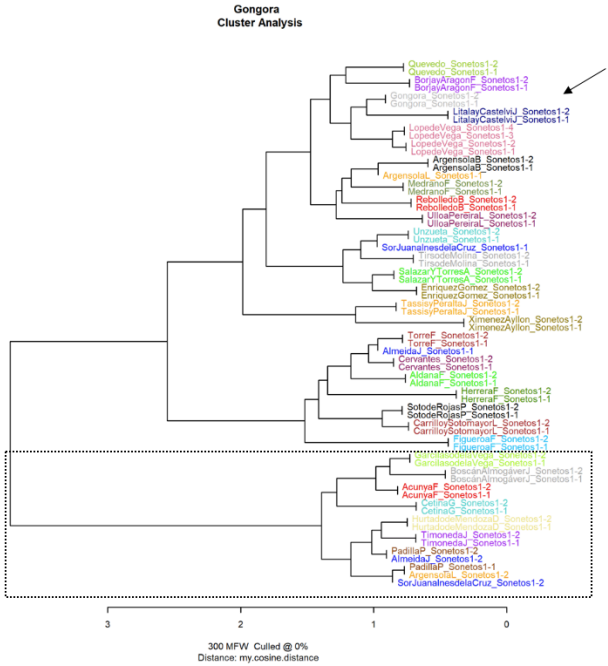



Figure 5: Dendrogram of the split sonnet-corpus (cf. fig. 4), only 300 MFW, arrow added to point out clustering of Góngora, Lope and Litala, colored box added to point out the arm of older sonnets

still used only 5 times out of approximately 1500 words.[33] Plus, we know that *stylo* takes into consideration the MFW of the whole corpus and here "pecho" is ranked on 85.[34]

All these experiments give us valuable insight about the literary history and a comparison of more authors than the usual canon. Proceeding with new experiments or following the surprisingly new path of Sor Juana Inés de la Cruz and the older Argensola brother would be another study. A result of the stylometric experiments until now is: Góngora is not the expected outlier

[33] For the calculation of words, I used *voyant-tools*: see for each corpus, Argensola (http://voyant-tools.org/?corpus=b96dd2492f53e3ed1a6d5c6663c32159) and Sor Juana (http://voyant-tools.org/?corpus=52bc60da5092b9ed8e707a8585a427d2). Sinclair, Stéfan and Geoffrey Rockwell, 2016. Voyant Tools. Web. http://voyant-tools.org/.

[34] To give an overview of the word list for the last experiment: first content word is "amor" (22.), followed by "vida" (43.), "sol" (45.), "alma" (48.), "cielo" (51.) – all of them could be used in every genre of sonnet: love, religion, satire, history.

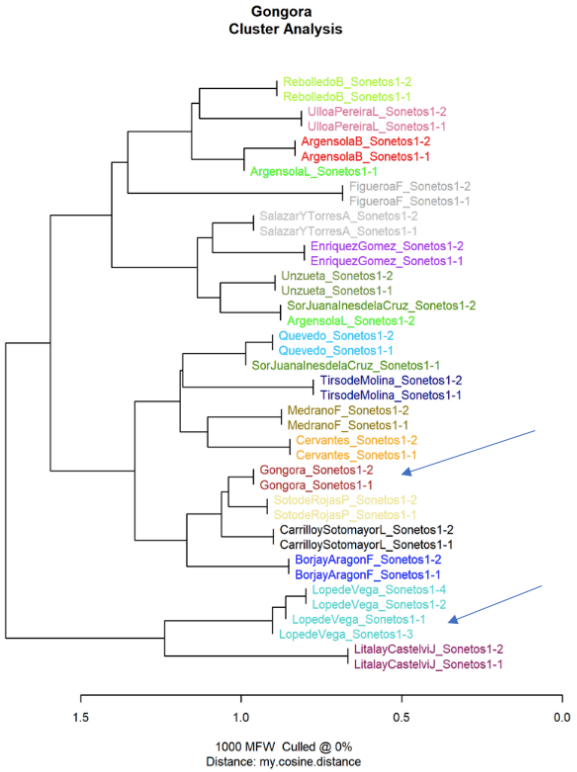



Figure 6: Dendrogram of the split sonnet-corpus, without 16th century-authors and Tassis y Peralta (less than 1000 words), arrows added to point out clustering of Góngora, Lope and Litala.

in the genre of sonnets and for contemporary authors. His role as a model and rebel for avantgarde literature is constructed by Lorca, the 27th generation and others[35] but defeated in recent studies, for example by Mercedes Blanco:

> La difficulté que nous éprouvons (ou devrions éprouver) à pointer tel trait poétique comme étant spécifiquement gongorin a en effet un caractère structurel : cette œuvre est tout le contraire d'une météorite ou d'un OVNI, et son

[35] See Federico Garcia Lorca, "La imagen poetica de Don Luis de Góngora", in *Obras completas*, Vol. 3 (Madrid: Aguilar, ²1986), 223–47; Hans-Ulrich Gumbrecht, "Warum gerade Gongora? Poetologie und historiales Bewusstsein in Spanien zwischen Jahrhundertwende und Bürgerkrieg" in *Lyrik und Malerei der Avantgarde*, ed. by Rainer Warning and Winfried Wehle (München: Fink, 1982), 145–92; Elsa Dehennin, *La résurgence de Góngora et la génération poétique de 1927* (Paris: Didier, 1962).

> auteur l'opposé d'un solitaire et d'un extravagant. Góngora, en phase avec son temps, à l'aise dans son monde, offre des solutions convaincantes bien qu'étonnantes à une recherche où tous sont embarqués, dans ce moment d'essor inégalé de la poésie espagnole qui se place entre les dernières décennies du XVI[e] et les premières du XVII[e] siècle.[36]

The presented stylometric experiments support the thesis of Blanco, that Góngora was not at all the bizarre author for which he was admired by the early Spanish avantgarde and other European modern authors like Picasso. Though, she does not contradict the assumption that Góngora tries to invent a new language. Instead, she tries to replace him in his period and points out his predecessors as well as his imitators.[37]

In the various recent studies on Góngora using digital methods or precisely stylometry we can see the new perspectives, but also its limits. As Blanco rightly points out, there exists no tool to explain the difficulty in understanding grammatical constructions like the hyperbaton or metaphors, poetic images, etc.

> En outre ce n'est pas par ce biais qu'il sera le plus facile de se servir des humanités numériques. Les logiciels de stylométrie sont pour l'heure basés sur le lexique et non sur la syntaxe, de sorte que le repérage et la quantification des entorses à l'ordre des mots standard (hyperbates), que l'on tient depuis toujours pour typiques du style de Góngora, semblent actuellement hors de la portée d'une observation automatisée.[38]

Here, we have to contradict her today by citing the work of her own scholar: Marie-Eglantine Lescasse is analyzing not only lexis but also syntax, using POS-Tagging combined with statistical tests and hermeneutic analysis.[39] Staying on the semantic level of lexis it is also possible to compare Góngora to his contemporaries. Antonio Rojas Castro presented in two recent studies the comparison of key-words in a corpus of 16-17[th] century *Fábulas* (fables) and for Góngora alone in three different periods (1580–1610, 1611–1618,

[36] Mercedes Blanco, "Sor Juana Inés et le programme Phœbus : tester sur le gongorisme un logiciel d'exploration de l'intertextualité", *e-Spania. Revue interdisciplinaire d'études hispaniques médiévales et modernes*, no. 29 (1. Feb. 2018), https://doi.org/10.4000/e-spania.27677.

[37] As the author of the study *Góngora, o, La invención de una lengua* Mercedes Blanco is trying to establish the grammatical and semantic new paths Góngora is taking as a model for other Spanish Golden Age authors. See for example her chapter on sonnets: 24–30.

[38] Blanco, "Sor Juana Inés et le programme".

[39] See Marie-Eglantine Lescasse, "Góngora hors norme? Étude stylométrique d'un motif gongorin", p. 91 in this volume.

1619–1625).[40] Arguing on the level of content words involves always more hermeneutic interpretation than arguing with the MFW. The conclusion, that out of the keywords it might be possible to define precisely the esthetics behind the poetry ("de identificar dos dominantes estéticas bien claras: por un lado, el registro cortesano; por el otro, el registro culto y sublime")[41] seems difficult, thinking about the various meanings of words in different contexts of poetry.[42] Though, it is not at all a "danger" of interpretation but a necessary part of stylometric analysis ("On remarque un saut périlleux entre les données qu'il obtient des logiciels et l'interprétation, séduisante, qu'il en propose.")[43]. At this point we are back on literary criticism, when out of the different methods, different results are discussed and interpreted. To participate in the debate of literary criticism, to combine and to compare stylistics and stylometry is one of the biggest challenge for digital Humanists, because there are two parties to convince: the one of statistical evidence (using the proper method?) and the one of hermeneutics (does the interpretation in the end have some impact in the literary debate?).

Stating that the results of stylometry contradict my hypothesis, that Góngora is an outlier of his period and at the same time support the thesis of Blanco and others, that Góngora is well established in his period, we have to change the research question: Góngora (and Picasso) try to invent a new language – yes, but at least Góngora (and probably Picasso too) are not alone on this project.

It is not possible to test this in the same way with Picasso's writings. Simply, because I could not find enough Spanish prose poems for the 20th cen-

[40] Antonio Rojas Castro, "Luis de Góngora y la fábula mitológica del Siglo de Oro: clasificación de textos y análisis léxico con métodos informáticos", *Studia Aurea* 11 (22. Dec 2017): 111–42, https://doi.org/10.5565/rev/studiaaurea.260; Antonio Rojas Castro, "¿Cuántos "Góngoras" podemos leer? Un análisis contrastivo de la poesía de Luis de Góngora", *e-Spania. Revue interdisciplinaire d'études hispaniques médiévales et modernes*, no. 29 (1. Feb. 2018), https://doi.org/10.4000/e-spania.27448.

[41] Rojas Castro, "¿Cuántos "Góngoras" podemos leer?".

[42] As Blanco and Navarro point out, it is probably more important to know hows the used words are ingeniously combined, than to know the words themselves, see Navarro, "A computational linguistic approach to Spanish Golden Age Sonnets", 112.

[43] Mercedes Blanco about the study of Rojas Castro in her article "Sor Juana Inés et le programme".

tury.[44] The experiments I made show rather an effect of genre clustering than of authorial style (fig. 7).

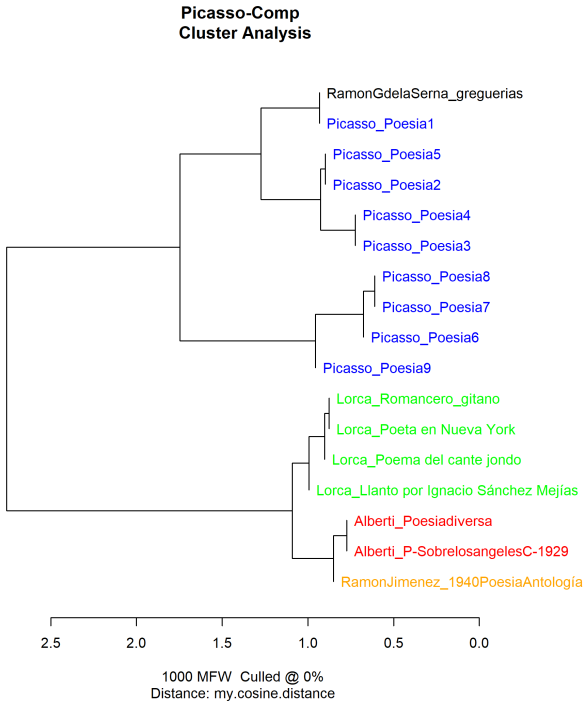



Figure 7: Dendrogram for the Picasso-corpus, generated with *stylo*

Here, I split the prose poems Picasso wrote in the years of 1935–1957 into 9 texts to have approximately the similar number of words as for the anthologies of poetry by Federico Garcia Lorca, Rafael Alberti, Juan Ramón Jiménez and Ramón Gómez de la Serna. Again, we can state that the attribution of authorship via *stylo* is correct and that prose (poems) and verse poetry is clustered together (genre signal). But, apart from those rudimentary observations every interpretation of style is not based on reliable statistic results. The

[44] There are anthologies of 20th century poetry in the BVMC, but most of them are verse poems, see: José Luis García Martín, ed., *Poetas del Novecientos: entre el Modernismo y la Vanguardia (Antología)*. Tomo I + II, edición digital basada en la de [Madrid, Fundación Santander Central Hispano, 2001] (Alicante: Biblioteca Virtual Miguel de Cervantes, 2004), http://www.cervantesvirtual.com/obra-visor/poetas-del-novecientos-entre-el-modernismo-y-la-vanguardia-antologia-tomo-i-de-fernando-fortun-a-rafael-porlan--0/html/.

number of authors compared is far too small for concluding anything about Picasso's standing in literary history. Plus, the genre difference between Picasso's prose poems, the modern verse poems of Lorca, Alberti and Ramón Jiménez and the *greguerías* by Gómez de la Serna influences the statistics far too much. In a direct hermeneutic comparison, we could prove similarities between Lorca, Alberti, Picasso and others.[45] But, they are due to topics, humour and for example the adoration for Góngora which is hard to measure in stylometry. Let me show with a very simple example how complex the relations via vocabulary and topics can be: Comparing three citations of Góngora, Picasso and Gómez de la Serna, we see that they are somehow talking about a swan, but all three in a completely different way and style:

| Góngora: *Soledades*, Soledad II, V. 799–805 (1618)[46] | Picasso: *Écrits*, 02.02.1936[47] | Gómez de la Serna: *Greguerías* (Madrid 1936)[48] |
|---|---|---|
| Can de lanas prolijo, que animoso buzo será, bien de profunda ría, bien de serena playa,<br>cuando la fulminada prisión caya del Neblí, a cuyo vuelo tan vecino a su cielo<br>el Cisne perdonara, luminoso | … del son del canto de cisne del clarín que se muere de pena nadando buzo ciego por las tripas verdes del lago por donde Ofelia va buscando en el fondo un trozo de papel y un lápiz… | La luna ilumina la cifra de almanaque de los cisnes.<br><br>Cuando el cielo se haya jugado todas las estrellas se jugará la Luna a cara o cruz.<br><br>En la Luna se han visto revoltear papeles de un picnic antediluviano. |

Table 5: Examples of Close Reading for Góngora, Picasso, Ramón Gómez de la Serna

The 'danger' of distant reading using stylometry and other digital methods is not at all the part of interpretation, but to be careful not to compare apples and oranges. In the previous experiments we have seen how much the composition of the corpus influences the performance of the clustering on the whole. Not until experiments with different parameters show the same effect, we can speak about reliable results. That is why I did not proceed with the stylometry on Picasso, because the corpus for comparison appears to be too poor for further analysis.

[45] See Rißler-Pipka, *Picassos schriftstellerisches Werk*, 238–60.

[46] Luis de Góngora y Argote, *Soledades*, Soledad II, V. 799–805, ed. by John Beverley (Madrid: Catedra, [14]2009), 158.

[47] Picasso, *Écrits*, 100–1.

[48] Ramón Gómez de la Serna, "Greguerías", *ABC (Madrid)* (22.07.1969): 96, http://hemeroteca.abc.es/nav/Navigate.exe/hemeroteca/madrid/abc/1969/07/22/096.html.

## Trying Topic Modeling on Picasso's writings in place of a conclusion

Topics are extremely important for Picasso's writings. Reading his texts from 1935 to 1957 the reappearing topics are probably the only stable and 'readable' in terms of 'understanding' that serves us as Ariadne's thread out of the labyrinth. Marie-Laure Bernadac tried even to make a dictionary or vocabulary for his writing.[49] But, to reduce and to explain Picasso's poetry by his topics simply cannot be a solution. It would deprive us from the pleasure of being lost in his word-labyrinth, of the humorous non-sense of his sequences (of words, numbers, signs), of the synesthetic perception and so on. Nearly every entry in Bernadac's dictionary is an ineligible reduction of sense:

> écriture […] L'écriture est au départ, en 1935, un recours quasi thérapeuthique qui lui permet de surmonter sa dépression […] éléments […] enfance […] ennui[50]

Difficult not to suspect an interpretation motivated one-sided by Picasso's biography as a celebrated artist and painter, reducing all the effort and literary power of his poetry to a psycho-analytic pathologic explanation. The raising question and simple falsification of this kind of interpretation is to detect Picasso's topics by using Topic Modeling on his writings without preliminary assumptions and without interpretation of the topic itself. The advantage of the method is its absolutely blindness regarding the person of the author (and his possible psychological problems). The disadvantage is that the methods and the model behind the method naturally assumes common texts to be analyzed. When John Firths explains distributional semantics as "a word is characterized by the company it keeps" – it is correct for the usual context and language.[51] But the poet, who intentionally changes the usual semantic context, probably destroys also the functionality of distributional semantics. Exactly this failure of the method can be an advantage for the literary scholar as Lisa Rhody points out:

> Topic modeling as a methodology, particularly in the case of non-figurative language texts like poetry, can help us to get to new questions and discoveries

[49] Marie-Laure Bernadac, "The Poetry of Picasso (with an Abridged Dictionary)", in *Écrits*, ed. by Christine Piot and Marie-Laure Bernadac (Paris: Gallimard, 1989), XIII-XXV.

[50] Bernadac, "The Poetry of Picasso", XIX.

[51] See for the reference of John Firth and an introduction into Topic Modeling using french examples: Christof Schöch, "Topic Modeling Genre: An Exploration of French Classical and Enlightenment Drama", *Digital Humanities Quarterly* 011, no. 2 (22. May 2017).

> – not because topic modeling works perfectly, but because poetry causes it to fail in ways that are potentially productive for literary scholars.[52]

That means, we are able to test two aspects by trying Topic Modeling on Picasso's writings: First, to show the interpretation of topics by Bernadac are rather driven by a biographical look on the texts than by an analysis of the text itself. Second, to show that Topic Modeling probably is going to fail as a method because it is created to detect topics in common text collections. A Topic is built by the statistical comparison of the use of words in a defined paragraph of text. That means all words of the corpus are compared to all other words. Then, statistics counts how often words appear together in the same paragraph (context). This is an extensive process because every single word has to be calculated in relation to every other word in the corpus and in relation to the distribution of words all over the corpus.[53] It is easy to imagine that words like "doctor, pharmacy, medicine, illness, temperature, etc". are more used together in one paragraph than mixed up with the word "landscape". In some corpuses these words belonging to the same lexical field are probably mixed with words like "child, night, car" or other terms that describe the situation in which the topic is used in this specific corpus.[54] Important to state though, that the results out of Topic Modeling are not necessarily the same as literary topics.

This test is far from being seriously using Topic Modeling as a method to really analyze topics represented in the text collection of Picasso. But, it is rather an 'experiment' to test if the algorithm 'reads' Picasso's poetry differ-

[52] Lisa M. Rhody, "Topic Modeling and Figurative Language", *Journal of Digital Humanities* Vol. 2, no. 1 (2012). http://journalofdigitalhumanities.org/2-1/topic-modeling-and-figurative-language-by-lisa-m-rhody/.

[53] For a very clear explanation of Topic Modeling for Spanish examples, see recently Borja Navarro-Colorado, "On Poetic Topic Modeling: Extracting Themes and Motifs From a Corpus of Spanish Poetry", *Frontiers in Digital Humanities* 5 (2018). https://doi.org/10.3389/fdigh.2018.00015.

[54] For prose and drama in most cases we can get sensible results, for example see the analysis by Ulrike Henny u. a., "Topic, Genre, Text: Topics im Textverlauf von Untergattungen des spanischen und hispanoamerikanischen Romans (1880–1930)", in *Jahrestagung des DHd-Verbands 2016: Modellierung, Vernetzung, Visualisierung. Die Digital Humanities als fächerübergreifendes Forschungsparadigma* (dhd2016, Leipzig, 2016), http://www.dhd2016.de/abstracts/vorträge-055.html; Nanette Rißler-Pipka und Christof Schöch, "Topic Modeling als Perspektive auf das spanische und französische Theater des 17. Jahrhunderts", in *Von Daten zu Erkenntnissen: Digitale Geisteswissenschaften als Mittler zwischen Information und Interpretation, Jahrestagung des DHd-Verbands* (dhd2015, Graz, 2015), 20, http://gams.uni-graz.at/o:dhd2015.v.015.

ently than a literary scholar like Bernadac (or myself). For a serious analysis I would have needed much more texts and probably a text collection for comparison. But, as a first step and thanks to the new Dariah Topics Explorer,[55] I tried my small collection of Picasso's Spanish texts.

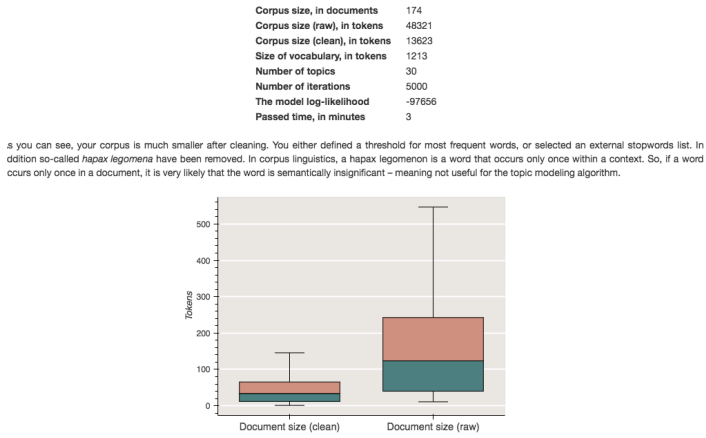



Figure 8: Screenshot: Result of DARIAH Topics Explorer: corpus and document size

One of the first results you get after using the tool is a description of the input (corpus), parameter and the difference between the document size with and without stop-words (functional words). Here, most of the words are functional words which is typical for prose, not necessarily for poetry. Further steps for deeper analysis could be the comparison to other corpuses (prose or poem or prose-poems) of the same period and language.

The next result, that is the most interesting, is the table of topics for the corpus (defined before by number, here: 30, see fig. 4). For comparison, I tried a well-prepared corpus of Hispanic novels and it seems easy to detect topics like "war" or "family", "rural life", "high society", "church", "virtue" etc. (fig. 9).

When I first tried the tool with the above given parameters and texts, the results were rather frustrating, because they did not much variate of the original texts. They did not form topics properly as shown for the corpus of hispanic novels of the 19th century (fig. 9):

[55] See Steffen Pielström et al., https://dariah-de.github.io/TopicsExplorer/.

| | Key 1 | Key 2 | Key 3 | Key 4 | Key 5 | Key 6 | Key 7 | Key 8 |
|---|---|---|---|---|---|---|---|---|
| **Topic 1** | paisano | partida | perro | compadre | toldo | cuero | especie | animal |
| **Topic 2** | fondo | pared | carne | papá | ocasión | valle | mozo | espalda |
| **Topic 3** | habitación | recuerdo | ángel | placer | sonrisa | sombra | presencia | rayo |
| **Topic 4** | mamá | chico | baile | negro | libro | patio | corredor | mozo |
| **Topic 5** | dama | convento | negro | fraile | cárcel | merced | justicia | palacio |
| **Topic 6** | ejército | tropa | guerra | ministro | emperador | comandante | presidente | oficial |
| **Topic 7** | virtud | cargo | naturaleza | do | concepto | término | impulso | intento |
| **Topic 8** | sociedad | salón | coche | fortuna | época | teatro | país | minuto |
| **Topic 9** | hacienda | cura | favor | juez | conde | interés | ropa | marqués |
| **Topic 10** | español | sacerdote | anciano | dios | piedra | país | bosque | rey |

Figure 9: Screenshot: Result of DARIAH Topics Explorer: Topics (here limited to 10) in the corpus of CLIGS-textbox, Hispanic novels of the 19th century[56]

Obviously, the Picasso-corpus was not that well prepared as the one cited above in comparison for Hispanic novels. We have a French text or parts of a French texts mixed up in the corpus (which is not that surprising because Picasso wrote some texts in both languages) (see Topic 6 in fig. 10). The elimination of stop-words did not work perfectly, because there are still words like "te", "tú", etc. Nevertheless, there are topics making sense in the context of Picasso's writings: for example, Topic 7 (corrida): "colores, caballo, abre, manos, toro, vientre, enamorado, bombo". But to be serious, we have to state, that the corrida-motif is spread out all over the topics in words like "traje" (for "traje de luces", the costume the torero wears), "aplausos" (in the arena de toros), even the innocent word "tarde" (afternoon) is strictly used in terms of the afternoon when the corrida begins. Looking at the topic of color and its lexical field it is even more striking that there is no real topic to identify (even Topic 11 contains only 3 color words), but it is spread all over the topics (see the highlighted words in fig. 10).

These, not very enlightening results, may lead to two conclusions. First, the tool did not work, because the parameters and input (not enough texts) were inappropriate. Or, second, that Picasso's style is refusing the rules of semantical distribution, because he refuses semantics at all. In Topic Modeling – as in most tools – there is always the danger that the observed effect is due to other reasons than we thought.

> In practice, collections are often constructed by combining documents from multiple sources, which may have distinctive style and vocabulary. This heterogeneity of sources leads to a serious but rarely studied problem: the

| | Key 1 | Key 2 | Key 3 | Key 4 | Key 5 | Key 6 | Key 7 | Key 8 |
|---|---|---|---|---|---|---|---|---|
| Topic 1 | color | cuchillo | tira | lleva | aun | ciego | necesario | sentado |
| Topic 2 | cara | oro | cuentas | día | sol | higos | puestas | tejado |
| Topic 3 | blanco | puesta | trapo | abajo | traje | sopa | punta | máquina |
| Topic 4 | sombra | mar | medio | lluvia | corazón | canta | deja | pico |
| Topic 5 | azul | plaza | gota | mañana | brazo | recoge | arriba | mira |
| Topic 6 | et | du | qui | sa | ce | dans | au | par |
| Topic 7 | colores | caballo | abre | manos | toro | vientre | enamorado | bombo |
| Topic 8 | color | leche | grano | lleno | oreja | peso | cal | cuelga |
| Topic 9 | tarde | medio | debajo | hechos | echando | cartero | niñas | caracoles |
| Topic 10 | canta | seda | vestido | tiempo | olor | jamón | puerta | sartén |
| Topic 11 | verde | azul | dedos | amarillo | medio | piel | alas | cielo |
| Topic 12 | aceite | pared | papel | agua | plata | estrellas | hilos | nubes |
| Topic 13 | silencio | mano | fondo | nido | verano | hora | canto | arde |
| Topic 14 | iris | cristal | oro | arco | aire | escalera | deseo | cuerpo |

Figure 10: Screenshot: Result of DARIAH Topics Explorer: Topics for Picasso-corpus as cited in fig. 8 – here cut by topic number 14 (of 30). Highlighted color words by N. R.-P.

> strongest, most prominent patterns in a collection may simply repeat the known structure of the corpus. Instead of finding informative, cross-cutting themes, models simply repeat the distinctive vocabulary of the individual sources.[57]

To find out if it is due to Picasso's individual distinctive style or a limitation of tokens, I increased the number of words artificially by quadruplicating Picasso's poems. Now the parameters are 20 topics, and 100 iterations (instead of 30 topics and 5000 iterations before) and the raw corpus consists of 209754 tokens (instead of 48321) (the corpus is cleaned up of French elements and stop words list is improved) (fig. 11)

Through the improved corpus the performance of the tool is obviously better. For example, the color theme is now concentrated in Topic 13 ("verde, azul, color…") and we find topics which consist of semantically linked words. In Topic 8 we have the food and bullfight combined: "sopa, cuchillo, harina" (soup, knife, flour) are cleary food and kitchen words and "tarde, traje" (afternoon, costume) belong to the bullfight in Picasso's vocabulary. Other examples of meaningful Topic-words are the rhyme words like "gritos, fritos" (screams, fried) in Topic 7 or "campana, ventana" (field, window) in Topic 10,

[57] Laure Thompson and David Mimno, "Authorless Topic Models: Biasing Models Away from Known Structure", (COLING 2018) http://www.cs.cornell.edu/~laurejt/papers/authorless-tms-2018.pdf.

| | Key 1 | Key 2 | Key 3 | Key 4 | Key 5 | Key 6 | Key 7 | Key 8 |
|---|---|---|---|---|---|---|---|---|
| Topic 1 | noche | alas | do | canta | cama | vestido | caballo | mil |
| Topic 2 | azul | detrás | manos | luz | sangre | sol | mano | hilo |
| Topic 3 | don | azul | medio | cura | puesto | tarde | digo | mar |
| Topic 4 | gori | caramelo | miel | tarde | quema | gotea | silencio | hora |
| Topic 5 | tiene | pájaro | recuerdo | jamón | pone | cuenta | pasar | equilibrio |
| Topic 6 | claro | día | verde | noche | después | silencio | salta | mesa |
| Topic 7 | noche | después | par | día | borde | seda | gritos | fritos |
| Topic 8 | tarde | patas | debajo | hechos | sopa | cuchillo | traje | harina |
| Topic 9 | tejado | cara | abre | ojos | toro | borde | amor | dedos |
| Topic 10 | papel | campana | ventana | piedra | oreja | agua | rosa | manos |
| Topic 11 | ala | cuarto | tantas | aroma | aire | blancura | tira | silencio |
| Topic 12 | sol | tío | luz | cuentas | cama | hora | ventana | vestido |
| Topic 13 | verde | azul | color | lleno | medio | cielo | amarillo | piel |
| Topic 14 | trigo | calor | trae | coche | azul | tarde | da | olor |
| Topic 15 | mano | cae | lleva | tira | cara | envuelto | cuchillo | necesario |
| Topic 16 | mano | lleva | cuerda | noche | reloj | hora | esconde | atraviesa |
| Topic 17 | tarde | cielo | monte | cristal | arco | aire | iris | oro |
| Topic 18 | sol | plumas | ruedas | negro | llena | olor | carne | pliegues |
| Topic 19 | blanco | pie | frío | plata | oro | colgando | cartero | pared |
| Topic 20 | gritos | mar | sartén | sol | lágrimas | tinta | par | medio |

Figure 11: Screenshot: Result of DARIAH Topics Explorer: Topics for Picasso-corpus (quadrupled), Highlighted boxes added by N. R.-P.

where also "peidra, oreja, agua, rosa" would follow the same rhyme, but less intensive only on the last letter. The preparation of the corpus is far from being perfect, there are still functional words like "debajo" (under) and "después" (then, later) and forms of verbs like "tiene" (has) or "pone" (put") which should be lemmatized and/or cleaned up to get a corpus of semantic words.

The next important question and experiment should be: How is Picasso's poetry performing in comparison to other poets of his tongue and time. To answer this question with the limited means I had, I took the same corpus as in stylometry for Picasso and added some of the same period, but with different style and poetics (Juan Ramón Jiménez, Ramón Gómez de la Serna, Rafael Alberti, Federico García Lorca, Jorge Guillen, Antonio Marchado) (see fig. 12 without Picasso and fig. 13 including Picasso).

For a comparison between Picasso and his contemporary Spanish colleagues, this is a simple method to point out differences in vocabulary. Adding Picasso in the experiment (fig. 13) shows that the vocabulary in the

| | Key 1 | Key 2 | Key 3 | Key 4 | Key 5 | Key 6 | Key 7 | Key 8 |
|---|---|---|---|---|---|---|---|---|
| Topic 1 | ángel | parece | ángeles | mar | alma | agua | tierra | viento |
| Topic 2 | tiempo | sol | azul | luz | vida | hombre | mar | mundo |
| Topic 3 | sangre | cielo | mundo | niño | ojos | luna | amor | caballo |
| Topic 4 | mujer | vida | parece | cielo | tarde | tiene | rosas | corazón |
| Topic 5 | poeta | tierra | abel | camino | mar | mañana | oh | mairena |
| Topic 6 | noche | luna | hombre | nieve | alto | lejos | frío | pena |
| Topic 7 | amargo | agua | verde | amor | tierra | noche | jinete | luna |
| Topic 8 | agua | tarde | corazón | sombra | tiene | tierra | piedra | ojos |
| Topic 9 | martín | amor | campo | día | mano | sol | alma | blanca |
| Topic 10 | mis | luz | mar | aire | ojos | sombra | alma | tus |

Figure 12: Screenshot: Result of DARIAH Topics Explorer: Topics for corpus of Spanish poetry 20th century (without Picasso), words in the corpus doubled, 10 Topics, 100 iterations. Highlighted boxes added by N. R.-P.

| | Key 1 | Key 2 | Key 3 | Key 4 | Key 5 | Key 6 | Key 7 | Key 8 |
|---|---|---|---|---|---|---|---|---|
| Topic 1 | alma | mar | sombra | viento | luna | ojos | mundo | piedra |
| Topic 2 | silencio | xxxv | sangre | aire | ventana | toro | noche | reloj |
| Topic 3 | día | mano | después | sol | cristal | decir | corazón | rosa |
| Topic 4 | par | tarde | cara | debajo | capa | medio | copa | sartén |
| Topic 5 | tierra | poeta | agua | martín | camino | oh | corazón | campo |
| Topic 6 | don | luz | azul | blanco | verdes | día | frío | ventana |
| Topic 7 | luz | luna | agua | noche | cielo | muerte | mujer | sangre |
| Topic 8 | verde | sol | azul | cielo | negro | piel | agua | alas |
| Topic 9 | tarde | amor | verde | vida | azul | sol | cielo | jardín |
| Topic 10 | azul | fuego | da | pan | color | colores | borde | olor |

Figure 13: Screenshot: Result of DARIAH Topics Explorer: Topics for corpus of Spanish poetry 20th century (with Picasso), words in the corpus doubled, 10 Topics, 100 iterations. Highlighted boxes added by N. R.-P.

whole corpus gets more disturbing, inconsistent, harder to understand. While in fig. 12 the first topic is clearly attributed to the angel poems by Alberti and looking into the distribution of topics in document would show where other poets are using this element. Another prominent and celebrated topic is the landscape and the new pilgrim in Machado's poetry which is here visible in Topic 5. Other topics are hinting to Lorca with the importance of "luz" (light) for example. But, adding Picasso's texts has the effect that not very well-defined topics out of the bullfight and color themes are mixed together (Topic 4 and 8). The other strong topics from the experiment before (fig. 12) are still visible (topic 5 with "tierra, poeta, …, camino, corazón, campo"), but it is interesting how Picasso uses some of the same words like "azul" (together with Lorca) to combine them with strange motifs like "tarde" (afternoon), "ventana" (window) or "pan" (bread). Before, in Topic 2 (fig. 12) "azul" was part of a transcendental reflection on life, man and the world ("tiempo, sol, azul, luz, vida, hombre, mar, mundo") which can be called a common topic for modern poetry.

Regarding the few tests, it can only be an attempt, concluding that Picasso is writing in a fragmented style, a montage of words, not in sentences, not in paragraphs, not in chapters, not in verses. Still after comparing Picasso's poetry to others, one can read the topics as a poem based on Picasso's box of words. The results are showing us the way Picasso writes, but they do not give us a one-sided explanation of the content. It helps us to understand the principle of montage – without filling the gaps between. For me, this effect is much more helpful and gives more insight than the hermeneutic explanation (for example by Bernadac). We can be curious about to hear what more can be done by trying algorithms on Picasso's poetry as Luis Meneses and Enrique Mallen presented on the subject at the DH2017 in Mexico.[58] Though, I would not agree, that the method could give us insights in how was "Picasso's 'vision' of the world".[59] There is no vision to be captured, but a way of creation that inspires and surprises his readers and spectators every time anew.

[58] Luis Meneses and Enrique Mallen, "Distinctions between Conceptual Domains in the Bilingual Poetry of Pablo Picasso". Mexico City, 2018. https://dh2018.adho.org/en/distinctions-between-conceptual-domains-in-the-bilingual-poetry-of-pablo-picasso/.

[59] Meneses and Mallen, "Distinctions".