



Transformers analyzing poetry: multilingual metrical pattern prediction with transformer-based language models

Javier de la Rosa¹ · Álvaro Pérez¹ · Mirella de Sisto¹ · Laura Hernández¹ · Aitor Díaz² · Salvador Ros² · Elena González-Blanco³

Received: 2 February 2021 / Accepted: 27 October 2021 / Published online: 15 November 2021
© The Author(s) 2021

Abstract

The splitting of words into stressed and unstressed syllables is the foundation for the scansion of poetry, a process that aims at determining the metrical pattern of a line of verse within a poem. Intricate language rules and their exceptions, as well as poetic licenses exerted by the authors, make calculating these patterns a nontrivial task. Some rhetorical devices shrink the metrical length, while others might extend it. This opens the door for interpretation and further complicates the creation of automated scansion algorithms useful for automatically analyzing corpora on a distant reading fashion. In this paper, we compare the automated metrical pattern identification systems available for Spanish, English, and German, against fine-tuned monolingual and multilingual language models trained on the same task. Despite being initially conceived as models suitable for semantic tasks, our results suggest that transformers-based models retain enough structural information to perform reasonably well for Spanish on a monolingual setting, and outperforms both for English and German when using a model trained on the three languages, showing evidence of the benefits of cross-lingual transfer between the languages.

Keywords Natural language processing · Language models · Digital humanities · Poetry

Mathematics Subject Classification 68T50 · 68U15

1 Introduction

In the last two decades, the almost coincidental emergence of the big data and distant reading [29] paradigms increased the demand within the Humanities for bigger corpora that could be analyzed in mass and from which trends and corpus-wide characteristics otherwise invisible could be identified. However, one particular literary genre that has remained somewhat overlooked by the advances in natural language processing is poetry. Different poetic

See <http://postdata.linhd.uned.es/>. Starting Grant research project Poetry Standardization and Linked Open Data: POSTDATA (ERC-2015-STG-679528) funded by the European Research Council (<https://erc.europa.eu>) (ERC) under the research and innovation program Horizon2020 of the European Union.

✉ Javier de la Rosa
versae@linhd.uned.es

Álvaro Pérez
alvaro.perez@linhd.uned.es

Mirella de Sisto
mdesisto@linhd.uned.es

Laura Hernández
laura.hernandez@linhd.uned.es

Aitor Díaz
adiazm@scc.uned.es

Salvador Ros
sros@scc.uned.es

Elena González-Blanco
egonzalezblanco@faculty.ie.edu

¹ LINHD, UNED, Juan del Rosal 16, Madrid 28040, Spain

² Control and Communication Systems, UNED, Madrid, Spain

³ School of Human Sciences and Technology, IE University, Madrid, Spain

traditions demand slightly different approaches to their analysis, despite being mostly based on the musicality and prosody of the specific languages they are rendered in. The analysis of poetry usually intertwines both structural and semantic aspects, making the creation of computer-based solutions a challenge. Extracting metrical patterns, calculating verse lengths, identifying rhyme schemes, or inquiring about rhyme types, are all parts of the study of poetry in structural terms. The scanning of a verse depends entirely on the correct assignment of stress to the syllables of the words is comprised of. This process might be affected by rhetorical figures and the particularities of each tradition. For example, one common device that can be found in Spanish, English, and German poetry is the synalepha, which allows to join separate phonological groups (syllables belonging to different words, i.e., the last syllable of a word and the first one of the next) into one single unit of pronunciation solely for metrical purposes (see Example 1). Two other such devices are syneresis, that operates similarly but within the word, and dieresis, that does the opposite by artificially splitting a syllable.

We can consider the metre of a verse as a sequence of stressed (strong) and unstressed (weak) syllables, which are sometimes denoted with the plus symbol ‘+’ for stressed syllables and the minus ‘−’ for the unstressed ones. While some traditions denote only the numeric positions of the stressed syllables, for clarity we will use the binary ‘+−’ codification in this study. Examples 1, 2, and 3 show verses of metrical lengths of 8, 10, and 7 syllables for Spanish, English, and German, respectively. These examples also show the resulting metrical pattern after applying synalepha (denoted by ‘~’) and considering the stress of the last word, which might also affect the metrical length in Spanish poetry.

Example 1 *cubra de nieve la hermosa cumbre*¹
cu-bra-de-nie-ve-la-her-mo-sa-cum-bre
 + − − + − − − + − + − 11
 (Garcilaso de la Vega)

Example 2 *Our foes to conquer on th’ embattled plain;*
Our-foes-to-con-quer-on-t ~ h’em-bat-tled-plain;
 − + − + − − − + − + 10
 (Rhys Prichard)

Example 3 *Leise lausch’ ich an der Thür*²
lei-se-la-sch ~ u’ich-an-der-Thür
 + − + − + − + 7
 (Adolf Schults)

Scanning poetry is deemed as a feasible task for humans while machines might struggle given the many rules and exceptions involved. Nevertheless, recent approaches to

natural language processing (NLP) have shown advantages over their rule-based counterparts. Modern NLP methods explore the idea of constructing numerical representations of documents (understood as sequences of words) that would allow to compare them using mathematical operations in a vector space. An especially powerful way of mapping a word into a numerical vector was introduced by Mikolov et al. [28]. Their word2vec algorithm, based on the distributional hypothesis, was able to find context-free vectors for words while retaining some of their semantic features. The approach was rapidly expanded to sentences [24] and documents [25]. However, one crucial aspect of natural languages, word polysemy, was left unaccounted for. Context-dependent word vectors tried to solve this issue by leveraging bidirectional long short-term memory neural networks and attention mechanisms that produced different vectors for a word depending on its context. Examples of these approaches are ULMFit [21], ELMo [32], or GPT [33], although it would be Devlin et al. [12] who would popularize multilingual language models with BERT. Research has shown that neural models implicitly encode linguistic features ranging from token labeling to different kinds of segmentation [26]. There is also evidence that language models and embeddings are able to capture not only semantic and syntactic properties but structural, as shown by Hewitt and Manning in their work with structural probes for extracting syntax trees [20], and Conneau et al. approximating the length in words of a sentence by its vector [11].

2 Related work

Approaches for the automated scansion of poetry date as far back as 1988 [27], at least for English. In this work, we only focus on recent and related advances for Spanish, English, and German as case studies. Our choice was guided on the importance of their poetic traditions as well as the availability of gold standard corpora and automatic metrical annotation tools for each language.

For Spanish, Gervás’ tool [15] is one of the earliest automatic annotation tools available. It uses definite clause grammars to model word syllabification and additional predicates to define synalepha, syllable count, and rhyme. More recently, the ADSO Scansion system introduced by Navarro-Colorado et al. [30] first applies part of speech (PoS) tags to the words of every line in a poem. The system can only handle verses of eleven syllables (hendecasyllables) and is capable of applying dieresis and synalephas as needed. Similarly, Rantanplan [34] employs PoS tags and syllabified words to assign stress. Unlike the ADSO Scansion system, Rantanplan applies all possible synalephas and syneresis at the syllable level before returning the

¹ “[It] covered with snow the beautiful summit.”

² “I quietly listen at the door”

metrical pattern. It is also currently the fastest and more accurate metrical annotation tool for Spanish poetry, and it works with different types of verses other than hendecasyllables. Agirrezabal et al. [1, 2] explored the idea of using recurrent neural networks bi-LSTM and CRF to automatically scan poetry in three languages (i.e. English, Spanish, and Basque). The tool tokenizes words, tags PoS, and assigns stress according to Groves et al. [16]. Their performance was not better than ADSO's nor Rantanplan.

Scandroid [19], first introduced in 1996, analyzes iambic and anapestic poetry, and served as an inspiration for many others similar tools for English. More recently, Antilla and Heuser [4] introduced Prosodic for metrical and phonological parsing. Its scansion process starts with tokenization of the text into words which are then converted into stressed syllabified phonetic transcriptions according to the CMU pronunciation dictionary. A metrical pattern is then assigned based on a set of customizable constraints. Built on Prosodic, Poesy [3] also detects rhyme patterns and groups syllables into feet. ZeuScansion [2] annotates poetry by defining line stress patterns and it also attempts to identify the dominant meter of a poem and which metrical feet constitute it. It uses tokenization, stress assignment via PoS tagging, and a pronunciation lexicon.

Finally, Metricalizer [5, 6] is a rule-based tool for metrical annotation of German poetry. The tool detects words and syllabifies them, and it is also capable of detecting lines and stanzas. The metrical annotation uses prosodic and morphological information. Rhyme recognition is based on the identification of vowels lengths, stressed syllables, and phonetic constituents. Also, metrical complexity is calculated by defining when metrical patterns diverge from prosodic structure. Other approaches for Middle High German exist [13, 14] but the differences with standard contemporary German are so profound that they cannot be reliably used to scan the same corpora.

3 Materials and methods

Given the encouraging previous results using transformer-based models and context-free embeddings for structural tasks, we decided to evaluate the capability of well performing language models to predict correct metrical patterns in the three languages. One challenging aspect of such comparison is the collection of the right annotated corpus.

As a corpus for Spanish, we decided to use the *Corpus de Sonetos de Siglo de Oro* "Golden Age Spanish corpus" [31]. This corpus, annotated in TEI-XML, contains sonnets from canonical Golden Age Spanish authors (16th and 17th centuries), featuring only hendecasyllabic verses. Although most of the poems included were annotated automatically,

it includes 730 poems with manually annotated metrical information, consisting of over 71,000 lines. From this corpus, a subset of 100 poems was used to evaluate ADSO Scansion system [30]. We also chose this subset as our test set (15%) and split the rest for training (70%) and evaluation (15%).

Unfortunately, for English and German we could not find annotated valid corpora of the scale found for Spanish. For English, while the Eighteenth-Century Poetry Archive (ECPA) [22] contains more than 3000 poems, at the time of writing around 95% of them seem to follow the same metrical pattern, thus making it useless for training purposes. Therefore, we chose an English corpus from *For Better For Verse* [35], an online platform of the University of Virginia for training students in annotating poetry. The 103 manually annotated poems composing the corpus are available in TEI-XML format. It was previously used in the literature for the evaluation of neural scansion systems for English [1]. For German, we used the manually annotated corpus from Haider and Kuhn [18] and Haider et al. [17]. The corpus contains 158 poems which cover the period from 1575 to 1936. Around 1200 lines have been annotated in terms of syllable stress, foot boundaries, caesuras and line main accent. The original non-annotated lines are available on the online platform *Antikoerperchen Lyrik Datenbank* "Little Antibodies Lyrics Database".³ Both of these corpora were also split in train, evaluation, and test sets following the same 70-15-15 rule applied for the Spanish corpus. Table 1 shows the number of lines in each split per corpus. While other corpora exist, they were not suitable for the task since the manually annotated metrical patterns were not varied enough or were simply missing.

4 Experimental design

With the available corpora, our downstream task is defined as metrical pattern prediction. That is, given a raw string of text representing a line of verse of a poem, a model is expected to predict a string of + and − symbols representing the stress of each syllable after any rhetorical device has been applied. Formally, it's a single-class multi-label classification task with as many labels as possible syllables in a verse. We defined two baselines based on fastText context-free embeddings of 300 dimensions with and without an extra BiLSTM before the prediction layer [7, 23]. We also selected the best performing methods for each language as the state-of-the-art in their respective languages. On the testing sets, we run the ADSO Scansion system [30] for Spanish, Poesy [3] for English, and Metricalizer [5] for German. ADSO and Poesy evaluations

³ <https://lyrik.antikoerperchen.de/>

Table 1 Number of verses for each language

	Train	Evaluation	Test
Spanish	7327	1421	1401
English	708	152	153
German	775	167	168

were run using a computer with an Intel® Core™ i7-8550U CPU @ 1.80GHz and 16GiB of DDR4 RAM memory. Metricalizer was run using their own online web-based tool.⁴

Using the training and evaluation sets, we fine-tuned several language models for each language and also for the three languages combined in different sets of experiments. We expected to see some gains in terms of cross-lingual transfer. Specifically, we used monolingual and multilingual BERT-base and RoBERTa models with a fully connected layer to predict the presence or absence of stress in each of the 11 positions of the hendecasyllabic verses in the Spanish corpora. The English and German corpora contained more varied verses in terms of metrical length, so we cut it at 12 and padded when needed. We used the language Python 3, the library PyTorch, and the framework Transformers [36] conveniently wrapped for classification tasks.⁵ We pre-processed all texts removing duplicated verses, lowercasing, and removing punctuation marks since they are irrelevant for metrical purposes. We found that lower numbers of epochs made the models perform very poorly. Therefore, we trained the models for 10 and 100 epochs using AdamW optimiser, warmup of 10%, and weight decay of 0.001. We used the evaluation set to search for the optimum learning rate between a set of $10e-6$, $15e-6$, $20e-6$, $30e-6$, $50e-6$, thus reporting on the best performing one. Training was done on a 8 vCPUs Google Cloud instance with 30GB of RAM memory and 4 NVIDIA Tesla V100 GPU with 16GB of memory running on Debian 10. The maximum sequence length was set at 24 tokens and the batch size for both training and evaluation was set to 8.

5 Results

It is commonplace in multi-label classification tasks to report on F-scores or even accuracy. However, in our case those metrics would produce per-syllable information distorting the results of our experiments. We decided to consider as a correct prediction only when all the

individual syllable predictors were correct in a per line basis. This is a much more strict and demanding requirement than it is usually needed, but for metrical purposes is an all or nothing: if a metrical pattern got one stressed syllable wrong, then the entire pattern is useless. As such, we are reporting accuracy expressed as a percentage of correct metrical patterns in the testing set.

The first thing we notice when looking at Table 2 is that our baselines performed very poorly. This suggests that the task at hand is not exactly trivial. There are marginal gains when increasing the number of epochs and applying Bi-LSTM layers on top of the context-free embeddings per language, but the accuracy is still far from state-of-the-art. Moreover, among all the monolingual BERT versions the only one outperforming the rule-based counterpart was English BERT-large [12] with a 38.82% accuracy, just a 0.66 percentage point increase over the English state-of-the-art. Although not shown in Table 2, English BERT-base performed on par to BERT-large. The Spanish [8] and German [9] BERT-base models improved with the number of epochs but remained far from their state-of-the-art scores. The multilingual version of BERT (mBERT) notably improved the scores of the monolingual versions for Spanish, performed better for English with fewer epochs, and yielded our best result for German (30.54%). The multilingual version of RoBERTa (XLM-RoBERTa) [10] only performed better than mBERT for Spanish. Interestingly, the best performing models for Spanish were the monolingual versions of RoBERTa [11]. Our guess is that despite being trained only on English data, the corpora used might contain enough Spanish words in their vocabularies to make Spanish downstream tasks feasible.

In order to test language transferability when applied to structural tasks, in a second set of experiments we decided to concatenate the datasets for the three languages. We then fine-tuned the models on the combined dataset and evaluated on the test sets for each individual language. Given the good performance of the supposedly English-only RoBERTa models, we decided to keep them in this set of experiments as well. As seen in Table 3, results for Spanish plateaued at 93.43%, suggesting we are reaching the limits of the dataset. On the other hand, we were able to outperform previous state-of-the-art for English with a 12.5% point increase in accuracy up to 50.66%, and a 3.6% point increase for German up to 48.50%.

⁴ <https://metricalizer.de/en/metrikanalyse/>.

⁵ <https://github.com/ThilinaRajapakse/simpletransformers>.

Table 2 Accuracy of the different methods fine-tuned and evaluated language by language on the test sets for 10 and 100 epochs. Best scores per language in bold, our best results underlined

Method	Spanish		English		German	
	10	100	10	100	10	100
Baseline (fasttext)	8.56	8.57	9.87	10.53	2.39	4.79
Baseline (fasttext + BiLSTM)	22.34	27.34	11.18	16.45	5.39	9.58
BERT (monolingual)	55.39	72.38	28.29	38.82	17.36	23.95
Multilingual BERT	87.72	90.44	32.89	35.71	20.96	<u>30.54</u>
RoBERTa (base)	89.86	93.22	22.37	32.89	23.35	28.74
RoBERTa (large)	89.36	<u>93.43</u>	26.32	30.92	23.95	28.74
XLM RoBERTa (base)	69.74	91.72	30.92	32.51	11.98	29.94
XLM RoBERTa (large)	83.30	92.29	32.89	32.89	10.78	29.94
SOTA	96.23		38.16		44.91	

Table 3 Accuracy of the different methods fine-tuned on the three languages for 10 and 100 epochs and evaluated independently on each language test set. Best scores per language in bold, our best results underlined

Method	Spanish		English		German	
	10	100	10	100	10	100
Multilingual BERT	87.29	90.01	35.53	41.45	37.72	39.52
RoBERTa (base)	76.52	87.37	36.18	36.21	31.74	43.11
RoBERTa (large)	87.29	<u>93.43</u>	30.92	40.79	29.94	42.51
XLM RoBERTa (base)	85.44	92.15	35.53	40.79	31.14	46.11
XLM RoBERTa (large)	82.44	93.29	29.31	50.66	34.73	48.50
SOTA	96.23		38.16		44.91	

6 Conclusions and further work

In this paper we have evaluated the capabilities of BERT-based models when trained on the task of predicting the metrical pattern of a verse. Under the assumption that transformed-based models were capable of performing tasks of structural nature beyond those of the semantic kind, we show that BERT models perform reasonably well for Spanish, while outperform the previous state-of-the-art for English and German.

Since the best performing models are those trained on a combined corpora, there is evidence of cross-lingual transfer in effect. This suggests that further training a specialized multilingual pre-trained model on poetic corpora could help improve on the task of metrical pattern prediction. Our result on language transferability paves the way for transformer-based multilingual models for metrical pattern prediction able to work on languages for which very few annotated corpora exist, as in the case of German. Traditionally, automated metrical pattern systems are built by hand for each individual language, which is a costly enterprise that could greatly benefit from using multilingual approaches like ours.

Moreover, our multilingual fine-tuned models could also assist in the creation of poems by analyzing the metrical structure of each verse generated by a third-party system.

Similarly, a whole variety of tasks could be also be tested: metrical length, enjambment detection, caesura detection and position, and synalephas, dieresis, and syneresis positions among others. It could also be interesting to apply domain-specific models at the stanza or even whole poem level to investigate whether BERT models could predict structure or poetic genre.

Acknowledgements This research was supported by the project Poetry Standardization and Linked Open Data (POSTDATA) (ERC-2015-STG-679528) obtained by Elena González-Blanco and funded by an European Research Council (<https://erc.europa.eu>) Starting Grant under the Horizon2020 Program of the European Union.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature. The study was conceived and designed by Javier de la Rosa. Material preparation, data collection and analysis were performed by Javier de la Rosa, Álvaro Pérez, Mirella de Sisto, Laura Hernández, and Aitor Díaz. The first draft of the manuscript was written by Javier de la Rosa. Salvador Ros commented on previous versions of the manuscript. Funding was provided by Elena González-Blanco. All authors read and approved the final manuscript.

Data availability statement The corpora and code used in this study are publicly available at the next code repository: <https://github.com/linhd-postdata/bertsification>

Declarations

Conflicts of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Agirrezabal M, Alegria I, Hulden M (2017) A comparison of feature-based and neural scansion of poetry. In: Proceedings of the international conference recent advances in natural language processing, Ranlp 2017, pp 18–23
- Agirrezabal M, Astigarraga A, Arrieta B, Hulden M (2016) Zeuscansion: a tool for scansion of english poetry. *J Lang Model* 4
- Algee-Hewitt M, Heuser R, Kraxenberger M, Porter J, Sensenbaugh J, Tackett J (2014) The stanford literary lab transhistorical poetry project phase II: metrical form. In: DH
- Anttila A, Heuser R (2016) Phonological and metrical variation across genres. In: Proceedings of the annual meetings on phonology, vol 3
- Bobenhausen K (2011) The metricalizer2-automated metrical markup of German poetry. *Current trends in metrical analysis*. Peter Lang, Bern, pp 119–131
- Bobenhausen K, Hammerich B (2015) Literary metrics, linguistic metrics, and the algorithmic analysis of German poetry using metricalizer (2). *Langages* 199:67
- Bojanowski P, Grave E, Joulin A, Mikolov T (2017) Enriching word vectors with subword information. *Trans Assoc Comput Linguist* 5:135–146
- Cañete J, Chaperon G, Fuentes R, Ho JH, Kang H, Pérez J (2020). Spanish pre-trained bert model and evaluation data. In: PML4DC at ICLR 2020
- Chan, B., Schweter, S., Möller, T (2020) German's next language model. In: Proceedings of the 28th International Conference on Computational Linguistics, pp. 6788–6796. International Committee on Computational Linguistics, Barcelona, Spain (Online) . <https://doi.org/10.18653/v1/2020.coling-main.598>. <https://www.aclweb.org/anthology/2020.coling-main.598>
- Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., Grave, E., Ott, M., Zettlemoyer, L., Stoyanov, V (2019) Unsupervised cross-lingual representation learning at scale. arXiv preprint [arXiv:1911.02116](https://arxiv.org/abs/1911.02116)
- Conneau, A., Kruszewski, G., Lample, G., Barrault, L., Baroni, M (2018) What you can cram into a single vector: Probing sentence embeddings for linguistic properties. arXiv preprint [arXiv:1805.01070](https://arxiv.org/abs/1805.01070)
- Devlin, J., Chang, M.W., Lee, K., Toutanova, K (2018) Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint [arXiv:1810.04805](https://arxiv.org/abs/1810.04805)
- Dimpel F (2015) Automatische mittelhochdeutsche metrik 2.0. *Phil. Netz* 73:1–26
- Estes A, Hench C (2016) Supervised machine learning for hybrid meter. In: Proceedings of the fifth workshop on computational linguistics for literature, pp 1–8
- Gervás P (2000). A logic programming application for the analysis of Spanish verse. In: International conference on computational logic, pp 1330–1344. Springer
- Groves PL (1998) Strange music: the metre of the English heroic line. *English Literary Studies* (University of Victoria)
- Haider, T., Eger, S., Kim, E., Klinger, R., Menninghaus, W (2020) Po-emo: conceptualization, annotation, and modeling of aesthetic emotions in German and English poetry. arXiv preprint [arXiv:2003.07723](https://arxiv.org/abs/2003.07723)
- Haider, T., Kuhn, J (2018) Supervised rhyme detection with Siamese recurrent networks. In: Proceedings of the second joint SIGHUM workshop on computational linguistics for cultural heritage, social sciences, humanities and literature, pp 81–86
- Hartman, C.O (2005) The Scandroid 1.1. <http://oak.conncoll.edu/cohar/Programs.htm>. Accessed 20 July 2020
- Hewitt J, Manning CD (2019) A structural probe for finding syntax in word representations. pp 4129–4138
- Howard, J., Ruder, S (2018) Universal language model fine-tuning for text classification. arXiv preprint [arXiv:1801.06146](https://arxiv.org/abs/1801.06146)
- Huber A (2020) Eighteenth-century poetry archive
- Joulin A, Grave E, Bojanowski P, Mikolov T (2016) Bag of tricks for efficient text classification. arXiv preprint [arXiv:1607.01759](https://arxiv.org/abs/1607.01759)
- Kim Y (2014) Convolutional neural networks for sentence classification. arXiv preprint [arXiv:1408.5882](https://arxiv.org/abs/1408.5882)
- Le Q, Mikolov T (2014) Distributed representations of sentences and documents. In: International conference on machine learning, pp 1188–1196
- Liu NF, Gardner M, Belinkov Y, Peters ME, Smith NA (2019) Linguistic knowledge and transferability of contextual representations. arXiv preprint [arXiv:1903.08855](https://arxiv.org/abs/1903.08855)
- Logan HM (1988) Computer analysis of sound and meter in poetry. *Coll Lit* 15(1):19–24
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S., Dean, J (2013) Distributed representations of words and phrases and their compositionality. In: Advances in neural information processing systems, pp 3111–3119
- Moretti F (2013) Distant reading. Verso Books, Brooklyn
- Navarro-Colorado B (2017) A metrical scansion system for fixed-metre Spanish poetry. *Digit Scholarsh Humanit* 33(1):112–127
- Navarro-Colorado B, Lafoz MR, Sánchez N (2016) Metrical annotation of a large corpus of spanish sonnets: representation, scansion and evaluation. In: International conference on language resources and evaluation, pp 4360–4364
- Peters, M.E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., Zettlemoyer, L (2018) Deep contextualized word representations. arXiv preprint [arXiv:1802.05365](https://arxiv.org/abs/1802.05365)
- Radford, A., Narasimhan, K., Salimans, T., Sutskever, I (2018) Improving language understanding by generative pre-training
- de la Rosa J, Pérez Á, Hernández L, Ros S, González-Blanco E (2020) Rantanplan, fast and accurate syllabification and scansion of Spanish poetry. *Procesamiento del Lenguaje Natural* 65:83–90
- Tucker HF (2011) Poetic data and the news from poems: a for better for verse memoir. *Vic Poet* 49(2):267–281
- Wolf T, Debut L, Sanh V, Chaumond J, Delangue C, Moi A, Cistac P, Rault T, Louf R, Funtowicz M et al (2019) Hugging-face's transformers: state-of-the-art natural language processing. ArXiv [arXiv:1910](https://arxiv.org/abs/1910)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.