Syllabification of the Divine Comedy

ANDREA ASPERTI, University of Bologna, Department of Informatics: Science and Engineering (DISI), Italy
STEFANO DAL BIANCO, University of Siena, Department of philology and literary criticism, Italy

We provide a syllabification algorithm for the Divine Comedy using techniques from probabilistic and constraint programming. We particularly focus on the *synalephe*, addressed in terms of the "propensity" of a word to take part in a synalephe with adjacent words. We jointly provide an online vocabulary containing, for each word, information about its syllabification, the location of the tonic accent, and the aforementioned synalephe propensity, on the left and right sides. The algorithm is intrinsically nondeterministic, producing different possible syllabifications for each verse, with different likelihoods; metric constraints relative to accents on the 10th, 4th and 6th syllables are used to further reduce the solution space. The most likely syllabification is hence returned as output. We believe that this work could be a major milestone for a lot of different investigations. From the point of view of digital humanities it opens new perspectives on computer assisted analysis of digital sources, comprising automated detection of anomalous and problematic cases, metric clustering of verses and their categorization, or more foundational investigations addressing e.g. the phonetic roles of consonants and vowels. From the point of view of text processing and deep learning, information about syllabification and the location of accents opens a wide range of exciting perspectives, from the possibility of automatic *learning* syllabification of words and verses, to the improvement of generative models, aware of metric issues, and more respectful of the expected musicality.

1 INTRODUCTION

The Divina Commedia is a famous narrative poem by Dante Alighieri, and one of the most influential works in European literature. It is structured into three *cantiche* (Inferno, Purgatorio and Paradiso), each one composed by 33 canti, plus an initial, introductory canto, traditionally considered as part of the Inferno.

The scheme of verses is the so called *terza rima*, based on lines of eleven syllables (hendecasyllables) [2] structured in *tercets* following the rhyme scheme aba, bcb, cdc, ded, …Each tercet is hence composed by 33 syllables, similarly to the number of canti in each cantica.

Differently from the quantitative verse of ancient classical poetry (Greek and Latin), the hendecasyllable of medieval and modern poetry is accentual. The key characteristic is the *stress on the tenth syllable*; in the frequent case where the final word is stressed on the penultimate syllable (parola piana), the previous constraint produces verses of eleven syllables. The verse also has a stress preceding the caesura [3, 6], on either the sixth or the fourth syllable. The poet may add additional accents providing rhythmic variations that help to obtain stylistic effects.

An essential prerequisite to follow the rhythm of the verse [4, 5] is to correctly identify its syllables. For several reasons that we shall explain in the following, this is not a trivial operation, at the point that one is frequently forced to rely on rhythmic constraints to understand the syllabification intended by the author.

The delicate point is the sequence of vowels, that in Italian, similarly to other European languages, can be either pronounced as a single sound (diphthongs/triphthongs) or as separate sounds (hiatus). The phenomenon of *dieresis* consists in modifying the pronunciation by reading a diphthong splitting the sounds as in a hiatus; conversely, we have a *syneresis* when a hiatus is joined into a single sound.

A similar phenomenon may happen between adjacent words, respectively ending and starting with vowels. In such a situation, we have to consider that the phono-syntactic chain in natural languages, including Italian, is continuous. When we speak, we don't make any pause between words. So two vowels facing each other from the borders of two different words normally behave just as if they belong to the same syllable. There is no hiatus. The metric formalization of this phenomenon is called synalephe. According to the structure of the language [8, 9, 24],

in Italian poetry synalephe is the norm, providing a sort of "default" way of reading a verse. However, there are many interesting exceptions, both in the perspective of the natural language and of the "poetic licence", which in Dante's case is supposed to be almost always well motivated, and aimed to obtain special effects according to the rhythm of the verse [12, 14, 20, 21]. The anomalous hiatus between two words is called *dialephe*.

For all this reasons, automatic syllabification of Italian poetry, and of Dante's work in particular, has been traditionally considered as a particularly challenging, almost hopeless task. We address the problem using techniques from probabilistic and constraint programming to generate the most likelihood syllabification of each verse, delegating to experts the solution of the most controversial cases. In the current version, we are left with a dozen problematic cases, discussed in Appendix C, whose complexity is not due to syllabification but to the their anomalous rhythm. The automatic identification of these verses is in fact one of the possible applications of our algorithm.

We believe this work may provide a base for a lot of interesting investigations. From the point of view of Neural Networks and Deep Learning, it would be interesting to see if we can teach an agent to correctly learn syllabification, and then use this additional knowledge to improve automatic generation of verses in Dante's style (we shall extensively discuss Poetry generation in Section 3). Imposing the knowledge of the syllabic structure on a character-base textual encoding could provide an essential insight on the role of vowels, consonants, and their mutual interplay. It would be particularly interesting if we could distinguish (groups of) vowels and consonants as different clusters in the latent space of their embedding.

From the linguistic point of view, we could provide automatic support for the identification of anomalous situations [12], the identification of periodic functions [7], or the classification of verses into different metric categories [11] (as well as their identification inside large corpora).

While our work is specific for Dante, the overall methodology can be easily adapted and fine tuned to any other author[1].

1.1 Achievements

The syllabification algorithm, developed in the Python language, is freely available on GitHub at the address https://github.com/asperti/Dante. The complete syllabification of the Comedy just takes a few seconds. The current release contains:

- full syllabification code;
- a copy of the Comedy's dictionary, saved as a pickle file and accessible as a python dictionary data structure. For each word we supply information relative to its syllabification (that in a few cases can be non deterministic), the position of the accent, and the synalephe propensity at left and right extremities;
- source version of the Divina Commedia, borrowed from the Gutenberg edition (see Section 2);
- syllabified version of the full Comedy, where syllables have been divided by a vertical bar as in

Nel |mez|zo |del |cam|min |di |no|stra |vi|ta,   Inferno I, 1

1.2 Structure of the work

The article is structured in the following way. In Section 2 we mention the digital source that we adopted as a base for our work, as well as other digital resources available on line, and their related projects. In Section 3, we discuss related works. Since we are not aware of any other similar syllabification effort, we provide a wider view on the state of the art in textual processing of poetry, with special emphasis on its generative modeling. In Section 4 we discuss the syllabification of words, that is intrinsically non deterministic. Section 5 is entirely devoted to our approach to the *synalephe*, that is the main contribution of this work. In Section 6, we discuss the structure of the *dictionary*, that provides, for each word, information about its syllabification, the location of its

---

[1]Some fine tuning seems to be unavoidable, since, e.g. the use of synalephe in Dante may be different from that of Ariosto.

accent and the synalephe propensity of the word with adjacent ones. In Section 7, we outline the syllabification algorithm, and our management of the metric constraints. Section 8 discusses some minor amendments that we did to the Gutenberg edition. A few problematic cases still remaining to be solved are reported in Section 9. The final Section 10 briefly summarizes the main contributions of the work, discusses possible improvements and hints to the many interesting research perspectives opened by this work.

A few addition material is provided in appendices. Appendix A contains, as an example, the full syllabification of the first Canto of the Inferno. Appendix B contains a detailed investigation of hiatus/diphthong situations in the Comedy. Appendix C provides the list of verses with anomalous rhythm currently identified by our algorithm.

2 DIGITAL SOURCE

A lot of different digital editions of the Divine Comedy are currently accessible on line; we mention a few of them at the end of this section. The edition we used for our project was edited by the Project Gutenberg Literary Archive Foundation; specifically, it is the ebook n. 1012 of 2015, that is a UTF-8 revision with special 8-bit characters of the version in the ebook n. 1000, based instead on the 7-bit ASCII character set. The Gutenberg source seems to essentially conform to Petrocchi's critical edition of the Comedy[2] [13], used as a reference by most of the digital editions available on line (see Section 2.1). The version provides a rich annotation with dieresis and stresses that help the correct syllabification of the text and the disambiguation of words. To make an example, the word Beatrice appears 64 times in the Comedy: in 43 cases the couple of vowels "ea" constitutes a diphthong, as in

> Io son Beatrice, che ti faccio andare,   Inferno II, 70

while in the remaining 21 cases it is a hiatus, as in

> tra Beatrice e te è questo muro.   Purgatorio XXVII, 36

In the above mentioned Gutenberg edition (as in Petrocchi's one), these hiatuses have been marked with a dieresis: Bëatrice. Not all hiatus/diphthong issues have been solved, however, as we shall see in Section 4.

As a matter of fact, our tool may provide a valuable assistance in the philological comparison of different editions; for instance we spotted at least one (plausible) error in the Gutenberg version, where there is a difference with Petrocchi's edition. It is interesting to observe that the mistake is shared by other digital versions on line, with a possible transfer between them: hence, it would have not been revealed by a mere textual comparison.

Apart from a few minor amendments listed in Section 8, we did a single modification to the Gutenberg text, consisting in replacing some occurrences of single quotes with double ones in a few quotations. The reason is that single quotes are identical to apostrophes, that as discussed in section 5.4, play a central role for syllabification; for this reason, it is better to avoid any confusion.

2.1 Other (re)sources

In this section, we mention just a few of the many important projects around the world providing digital editions of the Divine Comedy.

Dante Lab http://dantelab.dartmouth.edu/. This is an on-line database of Dante's commentators, from the 1320s to the present. It offers a virtual workspace supporting the simultaneous visualization and comparison of different texts. Dante Lab was supported by the Dartmouth Dante Project.

---

[2]As it is well known, no original manuscript of the Divine Comedy has survived, although there are hundreds of ancient manuscripts from the 14th and 15th centuries. The critical edition edited by Giorgio Petrocchi [13], published in four volumes between 1966 and 1967 as part of the National Edition of Dante's Works, is usually reputed to be the reference landmark for the Comedy.

Princeton Dante Project https://dante.princeton.edu/projinfo.html It offers a richly annotated electronic text for instructional and scholarly use. The digital source, freely available on line, is based on the critical edition by Giorgio Petrocchi [13].

Digital Dante https://digitaldante.columbia.edu/dante/divine-comedy/
The project results from a collaboration among the Department of Italian, Columbia University Libraries, and Columbia University Libraries' Humanities and History Division. One of its distinctive features is an original way to read and research intertextual passages in the Commedia called Intertextual Dante [25].

Dante Network https://www.dantenetwork.it/]
Dante Network is a platform hosted by the University of Pisa collecting data and tools for the investigation, the enrichment and the enhancement of works by Dante Alighieri. The most recent and innovative contribution is the Hypermedia Dante Network (HDN), which aims to extend to the Divina Commedia the ontology and the tools already tested on minor works of Dante Alighieri.

Wikisource Wikisource offers a couple of versions of the Divina Commedia; one based on Petrocchi's edition (https://it.wikisource.org/wiki/Divina_Commedia), and another one in the edition of Francesco da Buti[3] (https://it.wikisource.org/wiki/Commedia_(Buti) .

Enciclopedia Dantesca Treccani https://www.treccani.it/enciclopedia/elenco-opere/Enciclopedia_Dantesca
This enciclopedia, freely accessible on line, provides valuable information about Dante's vocabulary, and was a relevant source in our work.

3   RELATED WORKS

A pioneering application of computers to metric investigations of the Divina Commedia can be found in [7], aiming at a statistical analysis of the so called *periodic functions*, that is the identification of recurrent patterns in the position of accents inside the verse. The study, implemented in FORTRAN on a IBM 360/44, was performed on two subsets of 1272 verses (dataset A) and 1024 verses (dataset B) extracted according to different policies from the Comedy. The verses in these two subsets were *manually preprocessed* to split them into *metric units* of suitable length, as e.g.

$$1/4/2/2/2/$$

corresponding to accents at the following positions:

$$- + - - - + - + - + -$$

More recently, a similar investigation has been done by the second author in the case of the *Orlando Furioso* by Ludovico Ariosto [11].

However, at the best of our knowledge, and in spite of the many projects in digital humanities focused on Dante's work and the Divina Commedia, some of which have been recalled in Section 2.1, no one addressed so far the issue of automatic syllabification, nor provided as we do, a complete syllabification of the Comedy.

This is not at all trivial, due to the complex nature of the Italian hendecasyllable, and the sophisticated interplay between metric accents and synalephe. Since the objective of our study is original, our methodology and techniques are original too, and essentially developed from scratch.

Similarly to the field of digital humanities, there is a lot of interest, at present, on textual processing of poetic literature in the field of Deep Learning. Here, the final aim is the development of good generative textual models; there is a wide perception that the additional challenges posed by meter and rhyme could possibly drive the discovery of new techniques, fostering the development of the field. Since the seminal work of A.Karpathy [17], exploiting Recursive Neural Network for automatic textual generation of image captions, (see also his wondrous

---

[3]Francesco da Buti, also known as Francesco di Bartolo (Pisa or Buti, 1324 - Pisa, 25 July 1406), was an Italian literary critic and Latinist, and one of the first commentators on the Divine Comedy.

blog "The Unreasonable Effectiveness of Recurrent Neural Networks", where RNN are applied for the first time to Shakespeare's work) there has been a lot of work on Deep Learning techniques for Natural Language Processing (NLP), and also a specific interest on poetry, in many different languages, comprising e.g. English [19], Chinese [29], ancient Greek [18], and also Italian [30]. An additional source of complexity is that the problem can be addressed at different linguistic levels: characters, syllables, sub-words, or words. Moreover, the very notion of rhyme, in the Italian language, cover the last part of the word from the tonic accent, requiring knowledge about its position, and possibly ad-hoc tokenization. Works on poetry generation frequently rely on specific decoding procedures to generate reasonable poetry, involving selection from a set of candidate outputs [15, 23, 30]. Rhyming and metric constraints are somehow orthogonal to each other and can also be decoupled: working on the final word of verses, it is relatively easy to learn rhyming patterns relying e.g. on similarity matrix between words [16].

While rhyme is usually reputed to be the main problem of poetry generation, in the case of the Italian endecasyllable, metric constraints are the real challenge. As we shall explain, the musicality [10] of the verse is related to the position of stresses inside the verse, whose actual location, however, can be at some extent modified by interpretative modulations, via dieresis and dialephe. This is precisely what makes syllabification hard.

The first author recently proposed automatic generation of verses in Dante's style as a student project for his course of Deep Learning at the University of Bologna. In spite of the state-of-the art techniques exploited by students, comprising word embeddings [22], attention [1, 28], transformers [27] (replacing RNN in most NLP applications), including the "light" version (117 millions parameters) of GPT2 [26], results remained modest. The respect of endecasyllables is erratic, and understanding the rhyme structure is still problematic (jointly learning the *what* and the *where* still seems to be an issue). There are many justifications for these modest results, starting from the relatively small dimension of the training set. However, we believe that, in the case of Italian poetry, an additional source of problems is the lack of sufficient information in the source data, that is one of the motivations for decorating the text with the correct syllabification, and provide the position of the accent inside the word. This enriched information could also, at some extent, remedy to the limited dimension of the training set.

Independently from the generative task, our enriched format opens the way to a lot of interesting investigations, already hinted to in the introduction, especially in relation with a combined processing of the text at character, syllable, and word level, aimed to investigate their mutual roles.

4   SYLLABIFICATION OF WORDS

As we already recalled, the main problem of syllabification concerns the division of groups of vowels. Specifically, there are two main subcases, depending on whether the vowels are intra-word (hiatus/diphthong) or inter-word (synalephe). The synalephe is by far the most interesting and complex phenomenon, and the main subject of this research; it will be discussed in the next section. Our approach presupposes, among other things, a known, possibly non deterministic, syllabification of words, that we shall address in this Section.

The syllabification of single words, even if different from that of modern Italian, is not a really compelling problem; however, several words require ad-hoc treatment, and the situation is further complicated by the frequent occurrences of Latin words, and words that, in Dante's style, recall the spelling of Latin language, not to speak of the famous Provencal tercets of Arnaut Daniel (Arnaldo Daniello) in Purgatorio XIV, 140-147.

For all these reasons, we do not release a syllabification code, but instead provide a *full dictionary* with the syllabification of *each word* occurring in the Divine Comedy. This syllabification is *non-deterministic*, due to the fact that a single word may be split in different ways in different context, for metric reasons. A typical example, that we already discussed, is the word "Beatrice".

A deeper investigation of hiatuses and diphthongs in the Divine Comedy is given in Appendix B. In this section, we briefly discuss a few interesting issues, that may also help to understand the facilities offered by our tool for the analysis of the document.

As we already remarked, most of the hiatus/diphthong dichotomies are solved if the Gutenberg edition, by the use of dieresis. However,

(1) not all problems have been addressed in the text;
(2) the use of dieresis may be questionable.

Concerning the latter issue, we shall discuss a few delicate cases in Appendix B, mostly relative to the groups of vowels "io" and "ea".

Here we discuss a couple of the remaining hiatus/diphthong problems still present, in our opinion, in the Gutenberg edition. In all these cases, instead of superimposing notation on the text, e.g. in the form of additional dieresis, we exploit the intrinsic *non determinism* of our approach, allowing *multiple syllabifications* to draw upon in a probabilistic way. It is also important to observe that all these problematic samples have been *automatically* evinced by our algorithm, since syllabification failed or was extremely unlikely.

A first interesting case is relative to the word "creature" (creatures). Such word appears 8 times in the Divine Comedy; invariably, the sequence of vowels "ea" is a hiatus, but for a single unfortunate verse in the Paradiso:

> e queste cose pur furon creature; Paradiso VII, 127

Unfortunately, there is no accepted notation to express syneresis, so it is convenient to address the problem allowing a double, nondeterministic syllabification for the word "creature", with sensibly different probabilities.

Some other interesting cases are the words ending with -aio (primaio, Tegghiaio[4], migliaio and similar), which are normally spelt with a hiatus "a-io":

> E questi sette col primaio stuolo Purgatorio XXIX, 145
>
> migliaia di lunari hanno punita. Purgatorio XXII, 36

However, for an ancient habit in Italian poetry[5], the same suffix can appear as monosyllabic:

> Farinata e 'l Tegghiaio, che fuor sì degni Inferno VI, 79
>
> Quanto di qua per un migliaio si conta Purgatorio XIII, 22
>
> ne lo stato primaio non si rinselva. Purgatorio, canto XIV, 66

5 SYNALEPHE

The main problem for the correct syllabification of the Divine Comedy is the synalephe, that is the "melding" into a single syllable of two vowels belonging to two different, adjacent words [3, 5, 9, 21].

Our approach to the synalephe consists in associating to each extremity of a word a *probability*, that can be understood as its propensity for taking part into a synalephe. This probability can be computed statistically, computing the frequency of the phenomenon.

Since the other important metric information associated to a word is the number n of its syllables, each word can be schematically described as a triple

$$\langle p_l, n, p_r \rangle$$

where p_l and p_r are the synalephe probabilities (left, and right, respectively) and n is the number of syllables of the word (we will be forced to extend this basic representation with additional information).

---

[4]Tegghiaio Aldobrandi was mayor of San Gimignano and Arezzo; he fought in the battle of Montaperti (1260) as a Guelph.
[5]A habit which probably deals – by extension – with the monosyllabic treatment of the word "gioia" in Italian manuscripts, coming from the key-word "joi" in provencal poetry [21], pp. 293-4

The general idea is that the probability to have a synalephe between two adjacent words $w_1 = \langle p_l^1, n^1, p_r^1 \rangle$ and $w_2 = \langle p_l^2, n^2, p_r^2 \rangle$ is given by the product $p_l^1 p_l^2$.

For words starting/ending with a vowel, the (left/right) synalephe is the norm. So, in typical cases, the synalephe probability is 1 for vowels and 0 for consonants. For instance, the word "selva" is described by the triple $\langle 0, 2, 1 \rangle$, since it starts with a consonant, it has 2 syllables and it ends with a vowel; similarly, "oscura" is associated with the triple $\langle 1, 3, 1 \rangle$ since it starts and ends with a vowel, and it contains 3 syllables.

$$\text{selva} \qquad \text{oscura}$$
$$\langle 0, 2, 1 \rangle \quad \langle 1, 3, 1 \rangle$$

When we compose the two words together, the probability to have a synalephe is $p_r^1 p_l^2 = 1$. Since we have a synalephe, in the computation of the total number of syllables we need to subtract 1 to the sum of n_1 and n_2, In the above case, we have $2 + 3 - 1 = 4$.

When the probability of having a synalephe is neither 0 nor 1, we need to consider both possibilities, with the corresponding probabilities. In the end, we shall choose the most likelihood syllabification for the verse, compatible with the metric constraints. In principle, the number of cases may grow exponentially with the length of the verse. In practice, this is never a problem, since (a) relatively few words have not-categorical synalephe probabilities and (b) the length of the verse is in any case very small.

In our dictionary of the Divine Comedy we provide left and right estimations of the synalephe probabilities for each word.

The previous algorithm can be easily extended to take into account multiple, non deterministic syllabification of single words. For instance, we might be interested to associate to the word "avea" two possible representations, corresponding to the case of dipththong/hiatus (where the latter should have precedence on the former, if compatible with the metric):

$$\text{avea}$$
$$\swarrow \qquad \searrow$$
$$\langle 1, 2, 0.1 \rangle \qquad \langle 1, 3, 1 \rangle$$

In the case of diphthong, the words contains 2 syllables, and the right synalephe is unlikely; in the case of hiatus, the word contains 3 syllables, with definite propensity for synalephe. In the case "avea" is followed by another word starting with a consonant, the hiatus version will typically generate verses incompatible with the metric constraints, and the diphthong version will eventually prevail[6].

5.1 Metric constraints

The main metric constraint is to have a stress on the tenth syllable. As a rough approximation, we could expect 11 syllables for each verse, but a more precise investigation eventually requires to take into account the position of the stress inside the word. This is not a completely trivial operation in case of Dante, due to the old lexicon, the sometimes unusual conjugation of verbs and other problems. However, in our dictionary we also provide this auxiliary information. The "triple" of the previous section now becomes a quadruple

$$\langle p_l, n, a, p_r \rangle$$

where the new information a is an integer in the range [-(n-1),0] expressing the position of the accent as a negative offset from the right. For example, in "carità" the offset is 0, and in "Ettore" the offset is -2.

The main advantage of taking accents into consideration is the fact that, in addition to the stress on the tenth syllable, we can also add other metric constraints, for instance relative to the stress preceding the caesura,

---

[6]In the current version, we do not follow this approach for the word "avea". We systematically treat it as a diphthong, unless if otherwise required by the editor with a dieresis annotation.

on either the sixth or the fourth syllable (or more sophisticated patterns). Syllabifications not satisfying these constraints can be pruned or severely penalized in their likelihood.

There is just a slight difficulty in expressing the constraint relative to the stress on the tenth syllable, regarding what we may expect *after* the stress. If the stress is on the last word, we accept the verse even if it has more than 11 syllables, to take into account the few *versi sdruccioli* of the Comedy, such as:

che noi possiam ne l'altra bolgia scendere,	Inferno XXIII, 32
ch'era ronchioso, stretto e malagevole,	Inferno XXIV, 62
non da pirate, non da gente argolica.	Inferno XXVIII, 84

In general, we would be tempted to refuse any additional word after the one containing the stress in 10th, but this would rule out a few interesting cases of *composed rhyme* [5, 21]. For instance, in the following verse

e men d'un mezzo di traverso non ci ha	Inferno XXX, 87

the stress is on "non" and nevertheless it is followed by *two* additional words (counting as one, due to the synalephe between "ci" and "ha"). If you are curious, the verse is rhyming with "oncia" and "sconcia" (sic!).

Another example is

Poi che ciascuno fu tornato ne lo	Paradiso XI, 13

(ryhiming with "cielo" and "candelo"). The accent is on "ne" (a bland proposition, by the way), followed by an additional word.

The rule we adopt is that, after the word containing the accent on the tenth syllable (that may have arbitrary length), we may accept additional words provided they do not trespass the total amount of 11 syllables.

5.2 The role of punctuation

As in all known metric systems, the phono-syntactic continuity is not affected by punctuation marks. The presence of a punctuation mark is therefore substantially irrelevant for the purposes of the syllabification of the verse.

Normally, even the dot does not prevent synalephe:

per simil colpa». E più non fé parola.	Inferno VI, 57
fossero». Ed ei mi disse: «Il foco etterno	Inferno VIII, 73
perduto». Ed elli: «Vedi ch'a ciò penso».	Inferno XI, 15
sotto la pece?». E quelli: «I' mi partii,	Inferno XXII, 66
diss' io, «chi siete?». E quei piegaro i colli;	Inferno XXXII, 44

As a matter of fact, punctuation marks are an acquisition of modern philology, from the Renaissance onward, when the first publishers of printed texts faced the problem of how to reproduce the text of ancient manuscripts, where, as it is well known, punctuation is absent, or reduced to the essential.

5.3 Diving deeper into the synalephe

The idea of associating a synalephe probability 1 to vowels and 0 to consonants is of course just a basic, rough approximation. There are many exception to this rule and then exceptions to exceptions, leading us naturally to a probabilistic approach.

We now investigate a few subcategories of words deserving a special treatment.

Remark It is important to remark that the following discussion is *not* intended to be an exhaustive analysis of all cases of synalehpe/dialephe inside the Comedy. It is only meant to drive a reasonable *initialization* of

the synalephe probabilities, ruling out some relatively trivial cases. The algorithm will automatically spot the remaining problematic cases, for which probabilities can be assigned in an automatic or supervised way.

5.3.1 Words ending in an accented vowel. In the vast majority of cases, an accented final vowel requires dialephe. The words in the following, relatively short, list appear in the Comedy both with dialephe and synalephe (or exclusively with synalephe[7]):

apparì, bontà, ché, drizzò, fé, già, là, lì, lasciò, perché, però, più, portò, ricominciò, sé, sì, tornò, turbò

Here are some examples (the first one with synalephe, and the second one with dialephe)

portò
    sì che, stracciando, ne portò un lacerto.   Inferno XXII, 72
    A Minòs mi portò; e quelli attorse   Inferno XXVII, 124
perché
    perché appressando sé al suo disire,   Paradiso I, 7
    perché ardire e franchezza non hai,   Inferno II, 123
tornò
    Noi ci allegrammo, e tosto tornò in pianto;   Inferno XXVI, 136
    Così tornò, e più non volle udirmi.   Purgatorio XVI, 145

5.3.2 Words ending with a diphthong. In Dante, groups of vowels at the end of words are frequently treated as diphthongs/triphthongs. For instance:

    al pel del vermo r**eo** che 'l mondo fóra.   Inferno XXXIV, 108
    del gran dis**io**, di retro a quel condotto   Purgatorio IV, 9
    che pr**ia** m'av**ea** parlato, sorridendo   Paradiso XI, 17
    ch'uscir dov**ea** di l**ui** e de le rede;   Paradiso XII, 66

When a word with a potential hiatus is followed by another one starting with a vowel there is an ambiguous situation. For instance, in the verse

    tien alto lor disio e nol nasconde.   Purgatorio XXIV, 111

we could split "disio e" as "di|sio |e" or "di|si|o e".
Menichetti [21] treats these cases and calls them *diesinalefe*: when the possibility of a dieresis meets the possibility of a dialephe, the dialephe wins.

We privilege the diphthong also at the end of verses. So, a verse like

    mentre ch'io vissi, per lo gran disio   Purgatorio XI, 86

is going to be treated as an hendecasyllable with ten syllables.

This assumption is somewhat in contrast with the Italian metric tradition, where these cases are treated as *endecasillabi piani*, with eleven syllables [5, 21, 24].

However, the previous choice has essentially no impact on the metric structure of the verse, while it is sensibly simpler from the algorithmic point of view, since the syllabification of words becomes independent from their position inside the verse. If required, these cases are easily recognizable and could be simply fixed in a post processing phase.

In typical cases, and coherently with the previous politics, words ending with a diphthong have no propensity to synalephe.

---

[7]likely due to the small number of occurrences.

5.3.3   words starting with a diphthong. Similarly to the previous case, when a word start with "ia" (Iacopo, iaculi, iattura, …), "io" (Iosüè, Iove,…), "ie" (ier, iernotte, …), "iu" (iura, iube, Iunone, …) it is not inclined to (left) synalephe.

Examples are

«O Iacopo», dicea, «da Santo Andrea	Inferno XIII, 133
ché se chelidri, iaculi e faree	Inferno XXIV, 86
qual diverrebbe Iove, s'elli e Marte	Paradiso XXVII, 14
Chi dietro a iura e chi ad amforismi	Paradiso XI, 4

Other combinations of vowels are less evident (and less frequent, too). An interesting case is e.g. the word "uomini", sometimes participating in synalpehe

Li uomini poi che 'ntorno erano sparti	Inferno XX, 88

and sometimes not:

Ahi Genovesi, uomini diversi	Inferno XXXIII, 151

5.3.4   Monosyllables. Another class of words deserving a special attention is that of monosyllables. For instance, it is known [21] that a monosyllable followed by a word starting with an accented vowel typically results in a dialephe. However, in Dante, monosyllables frequently induce a dialephe even in different situations, like e.g.

a chi avesse quei lumi divini	Paradiso VIII, 25
dico che arrivammo ad una landa	Inferno XIV, 8
che alcuna virtù nostra comprenda,	Purgatorio IV, 2
punto del cerchio in che avanti s'era,	Paradiso XI, 14

It is also interesting to observe that many monosyllables end with with a diphthong (mio, tuo, suo, noi, voi, poi, cui, fui, sia, via, etc.), or are accented (già, ché, fé, già, là, più, sé, sì, etc.) so they also fall into the previous categories.

According to our investigation, the following additional monosyllables never occur in a synalephe in the Comedy:

Be, me, fa, fo, mo, Po, pro, qua, re, sto, te, tu, tra, tre

The following monosyllables deserves instead a probabilistic treatment (in addition to the accented words already listed in the previous section):

a, ad, che, chi, da, e, fra, fu, io, ho, ha, ma, o, qui, se, su, va   (*)

For instance, the quadruple associated with the conjunction "e" (and), is

$$\langle .9, 1, 0, .2 \rangle$$

meaning that it has a high propensity to take part in a synalephe on the left and is available (but not really inclined) to participate in a synalephe on the right. So, if there is the possibility to have a synalephe both on the right and on the left, the left one will be eventually preferred. For instance, in the verse

che membra feminine avieno e atto,	Inferno IX, 39

the syllabification would be "a|vie|no e|at|to" and not "a|vie|no| e at|to".

The synalephe probalities for the words in the (*) list are very different from each other. For instance, words like "da", "ma" or "fu" have very small synalephe propensity, that is typical of Dante; quoting Menichetti (translation by the authors):

> Dante always has hard attack (and dialephe) after the preposition "da". The scarce availability of a particle for elision is in many cases an indication that it is normally followed by a hiatus in the language, without having to be syntagmatically tonic: "Ma Amore", four syllables.
>
> <div align="right">Menichetti [21], p.338</div>

In general, the necessity of a probabilistic approach is readily understood in view of the following observation by Menichetti (translation by the authors):

> The unstressed vowels subject to apocope or elision ("di" better than "da", for example) and the initial apheretizable ones (especially i-) [...] bear particularly well the synalephe, and it is instead normal that they do more than other obstacles to the dialephe. Dante, who after the syntagmatically unstressed word "fu" almost always has dialephe (Par IX, 120), however, has synalephe with i- (Inf IXX, 63, Inf V, 54). Thus the dialephe of Par XX, 38 assumes special importance.
>
> <div align="right">Menichetti [21], p.339</div>

Remarkably, all verses mentioned by Menichetti are perfectly syllabified by our algorithm.

5.4   Apostrophes

An apostrophe is used to express the elision of a part of a word. Typically, the elision refers to a final or initial vowel, but in rare cases it may replace an entire syllable, as in "ver' ", "inver' " (towards), where the corresponding full words are "verso", "inverso", or also delete letters inside words, as in "acco'lo" (accoilo) or "entra'mi" (entraimi). These latter cases must be treated in a special way: in particular, the apostrophe has essentially no effect on syllabification.

In the other cases, the apostrophe should be basically treated as a vowel. For this reason, we prefer a syllabification of the form "do|v' or" over the more grammatical "dov' | or" (again, this can be possibly adjusted, if required, in postprocessing). Observe that, with our encoding of words, the two forms are sensibly different: "do|v' " has two syllables and is definitely prone to synalephe; "dov' " has a single syllable and refuses synalephe.

Our management of the apostrophe is however a bit more sophisticated. The point is that the apostrophe is not just expressing an elision, but frequently it is also meant to *induce* a synalephe in situations where, in principle, we would not expect one.

To make an example, let us consider the following verse:

> Così vid' i' adunar la bella scola   Inferno IV, 94

that must be syllabified as follows:

> Co|sì |vi|d' i' a|du|nar| la| bel|la| sco|la   Inferno IV, 94

The delicate point is the synalephe between "i'" and "adunar". In Dantes'work, the word "io" in similar situations *would not* generate a synalephe: consider for instance

> E io a lui: «L'angoscia che tu hai   Inferno VI, 43
>
> E io anima trista non son sola,   Inferno VI, 55
>
> per ch'io avante l'occhio intento sbarro.   Inferno VIII, 66

The following case is symmetric:

> per ch'io 'ndugiai al fine i buon sospiri,   Purgatorio IV, 132

As another example let us consider the verse

> E noi lasciammo lor così 'mpacciati.   Inferno XXII, 151

to be compared e.g. with

> che fu al dire e al far così intero.   Purgatorio XVII, 30

For all these reasons, we deserve to the apostrophe a special treatment: in our embedding of words as tuples, instead of the synalephe probability, we use the numerical value 2 to express the presence of an apostrophe (left or right, respectively): a sort of super-propensity to synalephe.

6 THE DICTIONARY

An important contribution of our work is to offer a digital dictionary of the Divine Comedy providing word level syllabification as well as the position of the stress inside the word. In addition, the dictionary gives left and right synalephe probabilities for each word, in the sense explained in Section 5.

More specifically, the information associated with each word w is a *list* of pairs (t, ws), where t is the metric-tuple $\langle p_l, n, a, p_r \rangle$ introduced in Section 5.1, and ws is the corresponding syllabification. We remember that n is the number of syllables in the word, a is the position of the accent (expressed as a negative offset from the right) and p_l, p_r are the synalephe probabilities for the word (left and right, respectively).

Usually the list associated with a word w is composed by a single entry, but in a few cases we need to take into account the possibility to have multiple, non-deterministic syllabifications.

The syllabification of words is expressed by a string where syllables are separated by a vertical bar character "|".

A typical usage of our algorithm consists in overriding an entry for a given word in the dictionary for discovering interesting cases of synalephe/dialephe (or also diphthong/hiatus) involving that word inside the Comedy. Most of the examples in this article have been obtained in that way.

7 THE VERSE SYLLABIFICATION ALGORITHM

We process a verse as a sequence of tokens $w_1, \ldots, w_n$. Tokens can be words or punctuation symbols; however, as explained in Section 5.2 punctuation symbols play essentially no role and can be neglected.

Remarkably, the only information we need for each word is its metric tuple $\langle p_l, n, a, p_r \rangle$.
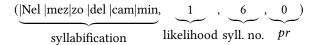
Suppose we already processed some initial part of the verse, up to token i. The information we have (the current *state*, in computer science terminology), is expressed in a list of possible syllabifications up the current token, where each syllabification is enriched with information regarding its likelihood p, the current number of syllables, and the synalephe probability of its last word.

To make an example, consider the simple verse

> Nel mezzo del cammin di nostra vita.   Inferno I, 1

and suppose that we already processed the initial part of the verse up to, say, the word "cammin". The current *state* would be described by the following tuple

$$(\underbrace{|\text{Nel} |\text{mez}|\text{zo} |\text{del} |\text{cam}|\text{min}}_{\text{syllabification}}, \quad \underbrace{1}_{\text{likelihood}} \quad , \quad \underbrace{6}_{\text{syll. no.}} \quad , \quad \underbrace{0}_{pr} \quad)$$

The first element is the actual syllabification; the second element is its likelihood, that in this case is 1, since it is deterministic (and hence unique), the third element is the current number of syllables, that in this case is 6, and the final one is the synalephe propensity of the last word "cammin", that is 0.

If we had multiple possible syllabifications, their likelihood would be distributed among them.

Suppose now we process the next token "di". We just need to take into account the possibility of a synalephe. In this case it is 0 because both adjacent words exclude it, so we add a syllable separator, concatenate the syllabification of the word "di" (obtained form the dictionary) to the current syllabification of the verse, update the

number of syllables, and remember the synalephe probability p_l of the last word "di". So, the new state will be:

$$(|Nel\ |mez|zo\ |del\ |cam|min\ |di, 1, 7, ?)$$

The probability of having a synalephe between two adjacent words with synalephe probabilities p_r and p_l (left and right, respectively) is just the product between the two probabilities, namely $p_r p_l$. In case one of p_r or p_l is 2 (corresponding to an apostrophe, see Section 5.4), than the resulting probability is 1, independently from the other.

If the probability p of having a synalephe is not categorical (0 or 1), we need to consider both possibilities, where the respective likelihood is the product between the current verse likelihood and the probability of the case under consideration (p for synalephe and $1 - p$ for absence of synalephe).

Let us discuss a slightly more complex example, considering the verse

esta selva selvaggia e aspra e forte   Inferno I, 5

We have no problems up to the first "e", where the state is

$$(|e|sta\ |sel|va\ |sel|vag|gia, 1, 7, 1)$$

Remember that the tuple for the word "e" is $\langle .9, 1, 0, .2$. So, after processing "e" we have two possibilities:

$$(|e|sta\ |sel|va\ |sel|vag|gia\ e, .9, 7, .2)$$
$$(|e|sta\ |sel|va\ |sel|vag|gia\ |e, .1, 8, .2)$$

The left synalephe probability for "aspra" is 1, so, for both states, the probability to have a synalephe is .2. After processing "aspra" we end up with 4 possible states, listed according to their probabilities

$$(|e|sta\ |sel|va\ |sel|vag|gia\ e\ |a|spra, .72, 9, 1)$$
$$(|e|sta\ |sel|va\ |sel|vag|gia\ e\ a|spra, .18, 8, 1)$$
$$(|e|sta\ |sel|va\ |sel|vag|gia\ |e\ |a|spra, .08, 10, 1)$$
$$(|e|sta\ |sel|va\ |sel|vag|gia\ |e\ a|spra, .02, 8, 1)$$

After processing the next "e" we get 8 possible states:

$$(|e|sta\ |sel|va\ |sel|vag|gia\ e\ |a|spra\ e, .648, 9, .2)$$
$$(|e|sta\ |sel|va\ |sel|vag|gia\ e\ a|spra\ e, .162, 8, 1)$$
$$(|e|sta\ |sel|va\ |sel|vag|gia\ e\ |a|spra\ |e, .072, 10, .2)$$
$$(|e|sta\ |sel|va\ |sel|vag|gia\ |e\ |a|spra\ e, .072, 10, .2)$$
$$(|e|sta\ |sel|va\ |sel|vag|gia\ e\ a|spra\ |e, .018, 9, .2)$$
$$(|e|sta\ |sel|va\ |sel|vag|gia\ |e\ a|spra\ e, .018, 9, .2)$$
$$(|e|sta\ |sel|va\ |sel|vag|gia\ |e\ |a|spra\ |e, .008, 11, .2)$$
$$(|e|sta\ |sel|va\ |sel|vag|gia\ |e\ a|spra\ |e, .002, 10, .2)$$

The processing of the last word "forte" does not introduce additional non determinism, so we end up with the 8 possible states listed below. However, only 3 of them are formed by 11 syllables, while the others (emphasized

in red) are not admissible:

(|e|sta |sel|va |sel|vag|gia e |a|spra e |for|te, .648, 11, .2)
(|e|sta |sel|va |sel|vag|gia e a|spra e |for|te, .162, 10, 1)
(|e|sta |sel|va |sel|vag|gia e |a|spra |e |for|te, .072, 12, .2)
(|e|sta |sel|va |sel|vag|gia |e |a|spra e |for|te, .072, 12, .2)
(|e|sta |sel|va |sel|vag|gia e a|spra |e |for|te, .018, 11, .2)
(|e|sta |sel|va |sel|vag|gia |e a|spra e |for|te, .018, 11, .2)
(|e|sta |sel|va |sel|vag|gia |e |a|spra |e |for|te, .008, 13, .2)
(|e|sta |sel|va |sel|vag|gia |e a|spra |e |for|te, .002, 12, .2)

Among the remaining possibilities with eleven syllables we choose the most likely, that is

|e|sta |sel|va |sel|vag|gia e |a|spra e |for|te

This choice is consistent with the prosody of Italian language, considering both the strong propensity of "e" to left synalephe ("selvaggia e", "aspra e"), and the propensity to natural hiatus between monosyllables and accented vowel ("e aspra").

7.1 Management of Metric constraints

For the reasons we already explained, instead of relying on the total number of syllables, it is better to rely on metric constraints. The most important one is to have a stress on the 10th syllable: if a syllabification does not satisfy this constraint it can be pruned. Additionally, we (currently) consider the constraint of having either a stress on the 6th or on the 4th syllable. More sophisticated constraints can be easily integrated in the algorithm.

In order to avoid to reprocess the entire verse, we add to each state a small list of boolean flags a4,a6,a10 expressing the presence of an accent at the corresponding position. The flag is initialized to False, and set to True if we find an accent on the expected syllable. In the current version, we accept accents independently from the grammatical category of the word (an information that we do not have in our vocabulary yet). We only reserve a special treatment to some monosyllables, mostly articles and prepositions, that we do not accept as legal accents.

The constraint on the tenth syllable is mandatory; for the other metric constraints, the algorithm privileges syllabifications satisfying either the stress in 4th or the stress in 6th, raising a warning if none of them is found. The verses raising a warning are just a dozen, and are commented in AppendixC; none of them is problematic from the point of view of the syllabification.

8 AMENDMENTS TO THE GUTENBERG EDITION

The verse

e suol di state talor essere grama.   Inferno XX, 81

contains 12 syllables. Following Petrocchi [13], we amend it as follows:

e suol di state talor esser grama.   Inferno XX, 81

It is interesting to observe that the previous mistake is common to several on line versions (at the time of submission), comprising e.g. https://divinacommedia.weebly.com/inferno-canto-xx.html, or https://www.hs-augsburg.de/ harsch/italica and many others.

8.1 Minor modifications

In the verse,

Tesëo combatter co' doppi petti;   Purgatorio XXIV, 123

we added a tonic accent on "combattér" (short for "combatterono"), to distinguish it from the infinite form "com-bàtter" in e.g.

licenza di combatter per lo seme   Paradiso XII, 95

In the verse,

ch'io drizzava spesso il viso in vano.   Purgatorio IX, 84

"ch'io" must be split into two syllables, while it is usually a single one. By analogy with similar situations in the document, we added a diereris: "ch'ïo".

9 PROBLEMATIC CASES

There still remain a few problematic (and debated) cases. The source of complexity is well described by Beccaria [3] in his discussion of the "dialephe" entry for the Enciclopedia Dantesca Treccani (translation by the authors):

> Indeed, Dante felt deeply in the verse, especially in the Comedy, the rhythmic and logical values (more than melodic fluidity). The dialephe, even of the less usual type in the Comedy, seems in fact very often determined by the pause of thought [...] and special artistic justifications were also found [12] (Casella) for slow narrative scans, underlined by dialephe before conjunctions.

It is remarkable to observe that most of the examples cited by Beccaria are automatically handled by our approach. For instance in the verse

d'infanti e di femmine e di viri   Inferno IV, 30

the dialephe after the word "infanti" (children) is imposed by the requirement of a stress on the 6th syllable.
Similarly, in

era già grande, e già eran tratti   Paradiso XVI, 107

the dialephe after "grande" (big) is required in order to have a stress on the 10th syllable. The most problematic cases are related to the heuristic rule of *detaching hemistics starting with an unstressed vowel in slow-scanned verses dealing with enumerations* [3].

We just mention the couple of cases that, according to this rule, are incorrectly classified by our algorithm (as far as we have been able to check); for these verses, we give the syllabification produced by our algorithm (former) and the "canonical" one (latter):

ma |sa|pï|en|za, a|mo|re| e |vir|tu|te   Inferno I, 104
ma |sa|pï|en|za, |a|mo|re e |vir|tu|te

leg|ge, |mo|ne|ta, of|fi|cio |e |co|stu|me   Purgatorio VI, 146
leg|ge, |mo|ne|ta, |of|fi|cio e |co|stu|me

Dealing with these remaining cases is an interesting challenge for future developments of our works. However, it is also worth to remark that it is precisely the somewhat anomalous behaviour of these verses that make them interesting from a linguistic point of view. So, the difference between the intended syllabification and the most likely one produced by our algorithm is *per se* an interesting phenomenon, worth to be observed and investigated more than simply corrected.

10 CONCLUSIONS

We give, for the first time, a complete syllabification of the Divine Comedy by Dante Alighieri. The interest of the work from the cultural heritage point of view is, in our opinion, evident: we enrich the source document

with an information that can be directly processed in a lot of different and innovative ways, and for different purposes. We enrich data, do not provide processing frameworks or applications.

Our syllabification algorithm, based on techniques borrowed from probabilistic and constraint programming, is simple and effective. It is not meant to mimic or reflect correct metric or phonetic rules: many of ours "admissible" syllabifications are totally incorrect from a metric point of view: we simply rule them out as unlikely. Our probabilistic approach could actually open new perspectives about the digital investigation of phonetic rules, and their empirical justification, especially in view of automatic learning.

While our work has been focused on the Divina Commedia, due to the world-wide interest on this document, the overall approach may be extended to most of the Italian poetic literature. The most delicate issue is the extension of the dictionary; in particular, it is not evident if the synalephe propensities generalize to different authors, and to what extent (that seems to be by itself a very interesting topic).

There are a few limitations of our approach, that can be possibly addressed in future works. A weak point, that could cause problems in some situations, is that the synalephe probabilities associated with words are *context independent*. Another interesting perspective, partially related to the previous point, consists in taking into account more sophisticated phonetic categories (semivowels, approximants, …). The management of punctuation symbols could be revisited in view of the issues in Section 9. Finally, as discussed in Section 7.1, the identification of the rhythmic cadence of a verse could take advantage of the grammatical category of words, that is an information currently missing from the vocabulary.

REFERENCES

[1] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate, 2014.

[2] Ignazio Baldelli. voce *endecasillabo*. In *Enciclopedia Dantesca*. Istituto della Enciclopedia italiana, Roma, 1970-78.

[3] Gian Luigi Beccaria. voci *cesura, dieresi, dialefe, ritmo*. In *Enciclopedia Dantesca*. Istituto della Enciclopedia italiana, Roma, 1970-78.

[4] Piero G. Beltrami. *Metrica, Poetica, Metrica Dantesca*. Pacini, Pisa, 1981.

[5] Piero G. Beltrami. *La metrica italiana*. Il Mulino, Bologna, 1984.

[6] Piero G. Beltrami. Cesura epica, lirica, italiana: riflessioni sull'endecasillabo di dante. *Metrica*, IV:67–107, 1986.

[7] Pier Marco Bertinetto. *Ritmo e modelli ritmici. Analisi computazionale delle funzioni periodiche nella versificazione dantesca*. Rosenberg & Sellier, Torino, 1973.

[8] Pier Marco Bertinetto. Strutture soprasegmentali e sistema metrico. *Metrica*, I:1–54, 1978.

[9] Pier Marco Bertinetto. *Strutture prosodiche dell'italiano. Accento, quantità, sillaba, giuntura, fondamenti metrici*. Accademia dell Crusca, Firenze, 1981.

[10] Dante Bianchi. Della "musicalità" considerata nella struttura del verso. *La Rassegna*, IV(33):81–113, 1925.

[11] Stefano Dal Bianco. *L'endecasillabo del «Furioso»*. Pacini, Pisa, 2007.

[12] Mario Casella. Studi sul testo della "divina commedia", ii. dieresi e dialefi d'eccezione. *Studi Danteschi*, VIII:28–63, 1924.

[13] Giorgio Petrocchi (editor). *La Commedia secondo l'antica vulgata, 4 voll.* A. Mondadori, 1966-67.

[14] Remo Fasani. *La metrica della «Divina Commedia» e altri saggi di metrica italiana*. Longo, Ravenna, 1992.

[15] Marjan Ghazvininejad, Yejin Choi, and Kevin Knight. Neural poetry translation. In Marilyn A. Walker, Heng Ji, and Amanda Stent, editors, *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT, New Orleans, Louisiana, USA, June 1-6, 2018, Volume 2 (Short Papers)*, pages 67–71. Association for Computational Linguistics, 2018.

[16] Harsh Jhamtani, Sanket Vaibhav Mehta, Jaime G. Carbonell, and Taylor Berg-Kirkpatrick. Learning rhyming constraints using structured adversaries. In Kentaro Inui, Jing Jiang, Vincent Ng, and Xiaojun Wan, editors, *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019*, pages 6024–6030. Association for Computational Linguistics, 2019.

[17] Andrej Karpathy and Fei-Fei Li. Deep visual-semantic alignments for generating image descriptions. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, Boston, MA, USA, June 7-12, 2015*, pages 3128–3137. IEEE Computer Society, 2015.

[18] A. Lamar and A. Chambers. Generating homeric poetry with deep neural networks. In *2019 First International Conference on Transdisciplinary AI (TransAI), Laguna Hills, CA, USA*, pages 68–75, 2019.

[19] Jey Han Lau, Trevor Cohn, Timothy Baldwin, Julian Brooke, and Adam Hammond. Deep-speare: A joint neural model of poetic language, meter and rhyme. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1948–1958, Melbourne, Australia, July 2018. Association for Computational Linguistics.

[20] Aldo Menichetti. Sulla figura di sinalefe / dialefe nel «canzoniere» di petrarca: l'incontro fra nessi bivocalici finali e vocale iniziale della parola seguente. *Studi petrarcheschi*, I:40–50, 1984.

[21] Aldo Menichetti. *Metrica italiana. Fondamenti metrici, prosodia, rima.* Antenore, Padova, 1993.

[22] Tomas Mikolov, Ilya Sutskever, Kai Chen, Gregory S. Corrado, and Jeffrey Dean. Distributed representations of words and phrases and their compositionality. In Christopher J. C. Burges, Léon Bottou, Zoubin Ghahramani, and Kilian Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. Proceedings of a meeting held December 5-8, 2013, Lake Tahoe, Nevada, United States*, pages 3111–3119, 2013.

[23] Hugo Gonçalo Oliveira. A survey on intelligent poetry generation: Languages, features, techniques, reutilisation and evaluation. In Jose Maria Alonso, Alberto Bugarín, and Ehud Reiter, editors, *Proceedings of the 10th International Conference on Natural Language Generation, INLG 2017, Santiago de Compostela, Spain, September 4-7, 2017*, pages 11–20. Association for Computational Linguistics, 2017.

[24] Mario Pazzaglia. *Teoria e analisi metrica.* Pàtron, Bologna, 1974.

[25] Julie Van Peteghem. *What is Intertextual Dante?* New York, NY: Columbia University Libraries., 2017.

[26] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. Language models are unsupervised multitask learners, 2019.

[27] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In Isabelle Guyon, Ulrike von Luxburg, Samy Bengio, Hanna M. Wallach, Rob Fergus, S. V. N. Vishwanathan, and Roman Garnett, editors, *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, 4-9 December 2017, Long Beach, CA, USA*, pages 5998–6008, 2017.

[28] Elena Voita, David Talbot, Fedor Moiseev, Rico Sennrich, and Ivan Titov. Analyzing multi-head self-attention: Specialized heads do the heavy lifting, the rest can be pruned. In Anna Korhonen, David R. Traum, and Lluís Màrquez, editors, *Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28- August 2, 2019, Volume 1: Long Papers*, pages 5797–5808. Association for Computational Linguistics, 2019.

[29] Zhuohan Xie, Jey Han Lau, and Trevor Cohn. From shakespeare to li-bai: Adapting a sonnet model to chinese poetry. In Meladel Mistica, Massimo Piccardi, and Andrew MacKinlay, editors, *Proceedings of the The 17th Annual Workshop of the Australasian Language Technology Association, ALTA 2019, Sydney, Australia, December 4-6, 2019*, pages 10–18. Australasian Language Technology Association, 2019.

[30] Andrea Zugarini, Stefano Melacci, and Marco Maggini. Neural poetry: Learning to generate poems using syllables. In Igor V. Tetko, Vera Kurková, Pavel Karpov, and Fabian J. Theis, editors, *Artificial Neural Networks and Machine Learning - ICANN 2019: Text and Time Series - 28th International Conference on Artificial Neural Networks, Munich, Germany, September 17-19, 2019, Proceedings, Part IV*, volume 11730 of *Lecture Notes in Computer Science*, pages 313–325. Springer, 2019.

A SYLLABIFICATION OF INFERNO, CANTO I

We recall our conventions:

- group of vowels in the final word of the verse are treated as diphthongs (tro|vai, in|trai, ab|ban|do|nai, …)
- apostrophes are assimilated to vowels: (on|d'io, do|v'or, …)

1	|Nel |mez|zo |del |cam|min |di |no|stra |vi|ta
2	|mi |ri|tro|vai |per |u|na |sel|va o|scu|ra,
3	|ché |la |di|rit|ta |via |e|ra |smar|ri|ta.
4	|Ahi |quan|to a |dir |qual |e|ra è |co|sa |du|ra
5	e|sta |sel|va |sel|vag|gia e |a|spra e |for|te
6	|che |nel |pen|sier |ri|no|va |la |pa|u|ra!
7	|Tan|t' è a|ma|ra |che |po|co |è |più |mor|te;
8	|ma |per |trat|tar |del |ben |ch' i' |vi |tro|vai,
9	|di|rò |de |l' al|tre |co|se |ch' i' |v' ho |scor|te.
10	|Io |non |so |ben |ri|dir |com' |i' |v' in|trai,
11	|tan|t' e|ra |pien |di |son|no |a |quel |pun|to
12	|che |la |ve|ra|ce |via |ab|ban|do|nai.
13	|Ma |poi |ch' i' |fui |al |piè |d' un |col|le |giun|to,
14	|là |do|ve |ter|mi|na|va |quel|la |val|le
15	|che |m' a|vea |di |pa|u|ra il |cor |com|pun|to,
16	|guar|dai |in |al|to e |vi|di |le |sue |spal|le
17	|ve|sti|te |già |de' |rag|gi |del |pia|ne|ta
18	|che |me|na |drit|to al|trui |per |o|gne |cal|le.
19	|Al|lor |fu |la |pa|u|ra un |po|co |que|ta,
20	|che |nel |la|go |del |cor |m' e|ra |du|ra|ta
21	|la |not|te |ch' i' |pas|sai |con |tan|ta |pie|ta.
22	|E |co|me |quei |che |con |le|na af|fan|na|ta,
23	|u|sci|to |fuor |del |pe|la|go a |la |ri|va,
24	|si |vol|ge a |l' ac|qua |pe|ri|glio|sa e |gua|ta,
25	|co|sì |l' a|ni|mo |mio, |ch' an|cor |fug|gi|va,
26	|si |vol|se a |re|tro a |ri|mi|rar |lo |pas|so
27	|che |non |la|sciò |già |mai |per|so|na |vi|va.
28	|Poi |ch' èi |po|sa|to un |po|co il |cor|po |las|so,
29	|ri|pre|si |via |per |la |piag|gia |di|ser|ta,
30	|sì |che |'l |piè |fer|mo |sem|pre e|ra |'l |più |bas|so.

31 |Ed |ec|co, |qua|si al |co|min|ciar |de |l' er|ta,
32 |u|na |lon|za |leg|ge|ra e |pre|sta |mol|to,
33 |che |di |pel |ma|co|la|to e|ra |co|ver|ta;

34 |e |non |mi |si |par|tia |di|nan|zi al |vol|to,
35 |an|zi ’m|pe|di|va |tan|to il |mio |cam|mi|no,
36 |ch’ i’ |fui |per |ri|tor|nar |più |vol|te |vòl|to.

37 |Tem|p’ e|ra |dal |prin|ci|pio |del |mat|ti|no,
38 |e ’l |sol |mon|ta|va ’n |sù |con |quel|le |stel|le
39 |ch’ e|ran |con |lui |quan|do |l’ a|mor |di|vi|no

40 |mos|se |di |pri|ma |quel|le |co|se |bel|le;
41 |sì |ch’ a |be|ne |spe|rar |m’ e|ra |ca|gio|ne
42 |di |quel|la |fie|ra a |la |ga|et|ta |pel|le

43 |l’ o|ra |del |tem|po e |la |dol|ce |sta|gio|ne;
44 |ma |non |sì |che |pa|u|ra |non |mi |des|se
45 |la |vi|sta |che |m’ ap|par|ve |d’ un |le|o|ne.

46 |Que|sti |pa|rea |che |con|tra |me |ve|nis|se
47 |con |la |te|st’ al|ta e |con |rab|bio|sa |fa|me,
48 |sì |che |pa|rea |che |l’ ae|re |ne |tre|mes|se.

49 |Ed |u|na |lu|pa, |che |di |tut|te |bra|me
50 |sem|bia|va |car|ca |ne |la |sua |ma|grez|za,
51 |e |mol|te |gen|ti |fé |già |vi|ver |gra|me,

52 |que|sta |mi |por|se |tan|to |di |gra|vez|za
53 |con |la |pa|u|ra |ch’ u|scia |di |sua |vi|sta,
54 |ch’ io |per|dei |la |spe|ran|za |de |l’ al|tez|za.

55 |E |qual |è |quei |che |vo|lon|tie|ri ac|qui|sta,
56 |e |giu|gne ’l |tem|po |che |per|der |lo |fa|ce,
57 |che ’n |tut|ti |suoi |pen|sier |pian|ge e |s’ at|tri|sta;

58 |tal |mi |fe|ce |la |be|stia |san|za |pa|ce,
59 |che, |ve|nen|do|mi ’n|con|tro, a |po|co a |po|co
60 |mi |ri|pi|gne|va |là |do|ve ’l |sol |ta|ce.

61 |Men|tre |ch’ i’ |ro|vi|na|va in |bas|so |lo|co,
62 |di|nan|zi a |li oc|chi |mi |si |fu |of|fer|to
63 |chi |per |lun|go |si|len|zio |pa|rea |fio|co.

64 |Quan|do |vi|di |co|stui |nel |gran |di|ser|to,
65 « |Mi|se|re|re |di |me», |gri|dai a |lui,
66 « |qual |che |tu |sii, |od |om|bra od |o|mo |cer|to!».

67 |Ri|spuo|se|mi:« |Non |o|mo, o|mo |già |fui,
68 e |li |pa|ren|ti |miei |fu|ron |lom|bar|di,
69 |man|to|a|ni |per |pa|trï|a am|be|dui.

70 |Nac|qui |sub |Iu|lio, an|cor |che |fos|se |tar|di,
71 e |vis|si a |Ro|ma |sot|to ’l |buo|no Au|gu|sto
72 |nel |tem|po |de |li |dèi |fal|si e |bu|giar|di.

73 |Po|e|ta |fui, |e |can|tai |di |quel |giu|sto
74 |fi|gliuol |d’ An|chi|se |che |ven|ne |di |Tro|ia,
75 |poi |che ’l |su|per|bo I|lï|ón |fu |com|bu|sto.

76 |Ma |tu |per|ché |ri|tor|ni a |tan|ta |no|ia?
77 |per|ché |non |sa|li il |di|let|to|so |mon|te
78 |ch’ è |prin|ci|pio e |ca|gion |di |tut|ta |gio|ia?».

79 « |Or |se’ |tu |quel |Vir|gi|lio e |quel|la |fon|te
80 |che |span|di |di |par|lar |sì |lar|go |fiu|me?»,
81 |ri|spuo|s’ io |lui |con |ver|go|gno|sa |fron|te.

82 « |O |de |li al|tri |po|e|ti o|no|re e |lu|me,
83 |va|glia|mi ’l |lun|go |stu|dio e ’l |gran|de a|mo|re
84 |che |m’ ha |fat|to |cer|car |lo |tuo |vo|lu|me.

85 |Tu |se’ |lo |mio |ma|e|stro e ’l |mio |au|to|re,
86 |tu |se’ |so|lo |co|lui |da |cu’ |io |tol|si
87 |lo |bel|lo |sti|lo |che |m’ ha |fat|to o|no|re.

88 |Ve|di |la |be|stia |per |cu’ |io |mi |vol|si;
89 |a|iu|ta|mi |da |lei, |fa|mo|so |sag|gio,
90 |ch’ el|la |mi |fa |tre|mar |le |ve|ne e i |pol|si».

91 « |A |te |con|vien |te|ne|re al|tro |vï|ag|gio»,
92 |ri|spuo|se, |poi |che |la|gri|mar |mi |vi|de,
93 « |se |vuo’ |cam|par |d’ e|sto |lo|co |sel|vag|gio;

94 |ché |que|sta |be|stia, |per |la |qual |tu |gri|de,
95 |non |la|scia al|trui |pas|sar |per |la |sua |via,
96 |ma |tan|to |lo ’m|pe|di|sce |che |l’ uc|ci|de;

97 |e |ha |na|tu|ra |sì |mal|va|gia e |ria,
98 |che |mai |non |em|pie |la |bra|mo|sa |vo|glia,
99 |e |do|po 'l |pa|sto ha |più |fa|me |che |pria.

100 |Mol|ti |son |li a|ni|ma|li a |cui |s' am|mo|glia,
101 |e |più |sa|ran|no an|co|ra, in|fin |che 'l |vel|tro
102 |ver|rà, |che |la |fa|rà |mo|rir |con |do|glia.

103 |Que|sti |non |ci|be|rà |ter|ra |né |pel|tro,
104 |ma |sa|pï|en|za, a|mo|re |e |vir|tu|te,
105 |e |sua |na|zion |sa|rà |tra |fel|tro e |fel|tro.

106 |Di |quel|la u|mi|le I|ta|lia |fia |sa|lu|te
107 |per |cui |mo|rì |la |ver|gi|ne |Cam|mil|la,
108 |Eu|ria|lo e |Tur|no e |Ni|so |di |fe|ru|te.

109 |Que|sti |la |cac|ce|rà |per |o|gne |vil|la,
110 |fin |che |l' av|rà |ri|mes|sa |ne |lo 'n|fer|no,
111 |là |on|de 'n|vi|dia |pri|ma |di|par|til|la.

112 |On|d' io |per |lo |tuo |me' |pen|so e |di|scer|no
113 |che |tu |mi |se|gui, e |io |sa|rò |tua |gui|da,
114 |e |trar|rot|ti |di |qui |per |lo|co et|ter|no;

115 |o|ve u|di|rai |le |di|spe|ra|te |stri|da,
116 |ve|drai |li an|ti|chi |spi|ri|ti |do|len|ti,
117 |ch' a |la |se|con|da |mor|te |cia|scun |gri|da;

118 |e |ve|de|rai |co|lor |che |son |con|ten|ti
119 |nel |fo|co, |per|ché |spe|ran |di |ve|ni|re
120 |quan|do |che |sia |a |le |be|a|te |gen|ti.

121 |A |le |quai |poi |se |tu |vor|rai |sa|li|re,
122 |a|ni|ma |fia |a |ciò |più |di |me |de|gna:
123 |con |lei |ti |la|sce|rò |nel |mio |par|ti|re;

124 |ché |quel|lo im|pe|ra|dor |che |là |sù |re|gna,
125 |per|ch' i' |fu' |ri|bel|lan|te a |la |sua |leg|ge,
126 |non |vuol |che 'n |sua |cit|tà |per |me |si |ve|gna.

127 |In |tut|te |par|ti im|pe|ra e |qui|vi |reg|ge;
128 |qui|vi è |la |sua |cit|tà |e |l' al|to |seg|gio:
129 |oh |fe|li|ce |co|lui |cu' |i|vi e|leg|ge!».

130	|E |io |a |lui:« |Po|e|ta, io |ti |ri|cheg|gio
131	|per |quel|lo |Dio |che |tu |non |co|no|sce|sti,
132	|ac|ciò |ch' io |fug|ga |que|sto |ma|le e |peg|gio,
133	|che |tu |mi |me|ni |là |do|v' or |di|ce|sti,
134	|sì |ch' io |veg|gia |la |por|ta |di |san |Pie|tro
135	|e |co|lor |cui |tu |fai |co|tan|to |me|sti».
136	|Al|lor |si |mos|se, e |io |li |ten|ni |die|tro.

B HIATUSES AND DIPHTHONGS IN THE DIVINE COMEDY

In this appendix, we provide a more exhaustive discussion about the syllabification of words in the Divine Comedy. The main purpose of this Section is to highlight the complexity of the problem and the differences with respect to the syllabification of modern Italian. The most interesting cases have been anticipated in Section 4; here we discuss a few additional issues regarding the use of hiatuses and diphthongs in the Divine Comedy. We organize the discussion around the most problematic combination of vowels: ia, ie, io, ea.

From a linguistic point of view, our discussion may actually appear over simplistic; actually, the presentation is mostly given from a computational perspective, and influenced by the current version of the algorithm. In particular, at the moment, we have no information about the position of the accent in a combination of vowels (this could be taken into account in future versions). In fact, in Italian poetry (and natural language) we can distinguish three situations:

(1) when the accent falls on the second vowel (ascendent nexus, es. patri-àrca, ri-òne, disi-àre) normally we have a hiatus (but never in the frequent cases of words coming in Florentine from the Latin PL: pieno from PLENUS, and in some other circumstances of real diphthongs [21]);

(2) when the accent falls on the first vowel (descendent nexus, es. farìa, ìo, mìo, arpìa, corsìa) the nexus is more often monosyllabic;

(3) when both vowels are out of accent (unstressed nexus, es. accìdia, ambròsia, viaggiàre, trionfàre) the situation is far more complicated, but in poetry prevails the hiatus [21].

B.1 i-a

The vowel "i" followed by "a", that typically constitutes a diphthong in modern Italian, is frequently a hiatus in the Divine Comedy (especially in the middle of words). A couple of frequent and relatively simple cases are the following:

viaggio
"A te convien tenere altro viaggio",   Inferno, I, 91

Other occurrences are in Inferno, X, 132; Inferno XVI, 27; Inferno XXVIII, 16; Inferno XXXI, 82; Purgatorio, II, 92.

disiata similarly to "disiato", "disiando", "disiar","disianza", etc.
Vostra parola disiata vola,   Purgatorio I, 83

Other occurrences are in Inferno V, 133; Inferno XXX, 140, Purgatorio III, 40, Purgatorio XXIX, 5; Purgatorio XXIX, 33; Purgatorio XXXIII, 83; Paradiso III, 73; , Paradiso V, 86; Paradiso XV, 66;; Paradiso XXII, 18; Paradiso XXII, 65; Paradiso XXIII, 4; Paradiso XXIII, 14; Paradiso XXIII, 39; Paradiso XXX, 15;

Other words that contain a ia-hiatus are:

accidia, ambrosia, Anfiarao, anzian, Briareo, eresiarche, celestiali, Ciriatto, clementiae, convertian, disia, disviando, disviato, Fialte, giovial, gloriar, gloriarla, gratia, Grazian, inebriate, India, inebriate, inebriava, 'nvetriate, inviasti, Iustiniano, Labia, lilia, Madian, mandrian, Mariza, Marsia, meridian, meridiana, Oriaco, Polinmia, potenziata, radial, radiando, Rialto, scienzia, scuriada, spezial, storiata, straniasse, sustanzial, svia, sviando, sviati, triangol, Trivia, umiliato, variar, variazion, venian, Vitaliano, Zodiaco

More complex is the case of the following words, sometimes pronounced with a hiatus, and sometimes with a diphthong. For each word, we provide an example of the two possibilities (we use a dieresis to remark the difference)

fiate,fiata

| | spesse fïate ragioniam del monte | Purgatorio XXII, 104 |
| | se mille fiate in sul capo mi tomi. | Inferno XXXII, 102 |

patria:

| | di quella nobil patrïa natio, | Inferno X, 26 |
| | e non molto distanti a la tua patria, | Paradiso XXI, 107 |

infamia

| | l'infamïa di Creti era distesa | Inferno XII, 12 |
| | sanza tema d'infamia ti rispondo. | Inferno XXVII, 66 |

venian

| | e non pareva, sì venïan lente. | Purgatorio III, 60 |
| | che venian lungo l'argine, e ciascuna | Inferno XV, 17 |

celestial

| | celestïal giacer, da l'altra parte, | Purgatorio XII, 29 |
| | Da poppa stava il celestial nocchiero, | Purgatorio II, 43 |

gloria

| | 'Glorïa in excelsis' tutti 'Deo' | Purgatorio XX, 136 |
| | de la mia gloria e del mio paradiso | Paradiso XV, 36 |

Although the sequence i-a frequently constitutes a hiatus, this is not a norm. In the frequent case in which the accent is on the second vowel, we have a natural diphthong. Examples, just borrowed from the first canto, are: pianeta, piaggia, piange, bestia, Troia, noia, gioia, Italia.

B.2 i-e

The two vowels "ie" form a hiatus in the following words:

Ariete, audienza, balbuziendo, coscienza, Daniel, Daniello, dieta, esperienza, esuriendo, Ezechiel, Gabriel, Gabriello, Galieno, 'nvieranno, niente, odierno, oriental, oriente, Ostiense, pazienza, pietate, progenie, quieta, quietar, quietarmi, quietata, quiete, quieto, quietò, requievi, riempion, riesca, sapienza, scienza, scienzia, Siestri .

The two vowels "ie" may be a hiatus in the following words:

pieta/pietate

| | con buona pïetate aiuta il mio! | Purgatorio V, 87 |
| | In te misericordia, in te pietate, | Paradiso XXXIII, 19 |

obediendo/disobediendo

| | con umiltate obedïendo poi | Paradiso VII, 99 |
| | quanto disobediendo intese ir suso; | Paradiso VII, 100 |

sufficiente/sufficienti

acciò che re sufficïente fosse;	Paradiso XIII, 96
per far l'uom sufficiente a rilevarsi	Paradiso VII, 116

Examples of natural dipththongs from the first Canto are: pensier, pien, fiera, volontieri, miei, convien, empie, Pietro, dietro.

B.3 i-o

The two vowels "io" are a hiatus in the following words:

accidioso, Anfione, anterior, aspersion, Caliopè, Curio, Diogenès, Diomede, Dione, Dionisio, disio, disioso, division, elezion, elezione, elezioni, Eliodoro, Etiopia, Etiopo, furiosa, gaudiose, gaudioso, Gerion, Gerione, gloriosa, gloriosamente, glorioso, gloriosi, idioma, Iperione,lioncel, Livio, lussuriosa, Niobè, oblivion, oppinion, oppinione, Pigmalione, pio, piorno, presunzion, preziosa, prezioso, razionabile, region, religione, Scariotto, Scipion, Scipione, settentrion, settentrional, studiose, violenta, violenti, viole, violenza, vision, visione,

The two vowels "io" may be a hiatus in the following words:

conversione

La mia conversïone, omè!, fu tarda;	Purgatorio XIX, 106
e per trovare a conversione acerba	Paradiso XI, 103

distinzione

sanza distinzïone in essordire.	Paradiso XXIX, 30
che sanza distinzione afferma e nega	Paradiso XIII, 116

grazioso

O animal grazïoso e benigno	Inferno V, 88
ditemi, ché mi fia grazioso e caro,	Purgatorio XIII, 91

invidiosa/invidiosi

silogizzò invidïosi veri.	Paradiso X, 138
gent' è avara, invidiosa e superba:	Inferno XV, 68

orazion/orazione

se buona orazïon lui non aita,	Purgatorio XI, 130
così, a l'orazion pronta e divota,	Paradiso XIV, 22

passion

quand' ira o altra passïon ti tocca!	Inferno XXXI, 72
a la passion di che ciascun si spicca,	Purgatorio XXI, 107

perfezion/perfezione

di tutta l'animal perfezïone;	Paradiso XIII, 83
sanza sua perfezion fosser cotanto.	Paradiso XXIX, 45

The cases of the two words "mio" (my) and "io" (I) possibly deserves a little discussion.

The word "mio" occurs 310 times in the Divine Comedy, and in 307 cases it must be read as a single syllable. Let us briefly review the remaining three cases, namely

ma quella folgorò nel mïo sguardo	Paradiso III, 128
Tal vero a l'intelletto mïo sterne	Paradiso XXVI, 37
già tutta mïo sguardo avea compresa,	Paradiso XXXI, 53

These verses are discussed in [21], p.250, who remarks the relation with the spurious s- following "mïo". Actually, the ancient Florentine language still maintains the *prostesi* in many cases ("iscritto", "isguardo", . . .); the potential *prostesi*, helps the separation of the previous nexus.

The case of "io" (I) is even more problematic. In the Gutenberg edition, we have 28 occurrences of "ïo", versus 679 occurrences of "io". Due to the synalephe, in most of the cases the two forms "io/ïo" would not change neither the total number of syllables, nor the position of tonic stresses inside the verse. The hiatus "ïo" may look closer to the Latin etymology "ego", but in the frequent cases when synalephe is impossible, it is clear that "io" normally constitutes a single syllable. Some examples are:

rispuos' io lui con vergognosa fronte.	Inferno I, 81
tu se' solo colui da cu' io tolsi	Inferno I, 86
Io non posso ritrar di tutti a pieno,	Inferno IV, 145
e di questi cotai son io medesmo.	Inferno IV, 39
Io venni in loco d'ogne luce muto,	Inferno V, 28

Nevertheless, there are situations where a hiatus is imposed for metric reasons. A nice example is the following verse, where the two forms coexists:

Cred' ïo ch'ei credette ch'io credesse	Inferno XIII, 25

Let us also observe that shifting the hiatus on the second occurrence of "io" would not change the total number of syllables. However, the hiatus on the first occurrence allows us to have a stress on the 6th syllable, in "credètte", and is eventually correct. This is a clear indication that for the correct syllabification of verses relying on the stress in 10th is not enough.

The linguistic motivation to privilege the dieresis on the first occurrence of "io" is that, coming after its verb, it is under stress and produces a more natural dieresis [21].

B.4   e-a

The group "ea" is the one where the use of dieresis in the Gutenberg edition is more questionable.

The word "ideale" is expressed with dieresis, but this looks somehow redundant.

Apart from a few proper names (Beatrice, Enea, Rea, Rodopea, Tarpea, …) the remaining cases are verbs at the imperfect:

avea, avean, correan, dicea, discendea, discernea, dovea, facea, facean, giacea, imprendea, intendea, parea, parean, percoteansi, piacea, piangea, potea, potean, ravvolgea, reflettea, sapean, solea, tenea, vedea, vincea.

In all these verbs, "ea" may constitute both a hiatus or a diphthong (where the latter is, by far, the norm). We give a few examples

avea

fiso nel punto che m'avëa vinto.	Paradiso XXIX, 9
ma l'un de' cigli un colpo avea diviso.	Purgatorio III, 108

avean

udito avëan l'ultimo costrutto;	Purgatorio XXVIII, 147
Non avean penne, ma di vispistrello	Inferno XXXIV, 49

correa/correan

corrëan genti nude e spaventate,	Inferno XXIV, 92
e correa contro 'l ciel per quelle strade	Purgatorio XVIII, 79

dovea

Sì com' io fui, com' io dovëa, seco,	Purgatorio XXXIII, 22
lo qual dovea Penelopè far lieta,	Inferno XXVI, 96

facea

ch'io facëa dinanzi a la risposta,	Inferno X, 71
sì che 'l sangue facea la faccia sozza,	Inferno XXVIII, 105

parean

e or parëan da la bianca tratte,	Purgatorio XXIX, 12
Morti li morti e i vivi parean vivi:	Purgatorio XII, 67

piangea

Io non piangëa, sì dentro impetrai:	Inferno XXXIII, 49
quando piangea, vi facea far le grida.	Inferno XIV, 102

However, there are a few cases where Petrocchi's use of dieresis looks somehow abused. Consider for instance the following verses:

Ella non ci dicëa alcuna cosa,	Purgatorio VI, 64
che 'l cibo ne solëa essere addotto	Purgatorio XXXIII, 44

As a matter of fact, there is no reason to expect synalephe in these cases (and hence no reason add a dieresis), precisely *because* the words on the left end with a diphthong. This is a case of the already mentioned *diesinalefe* rule by Menichetti: when the possibility of a dieresis meets the possibility of a dialephe, the dialephe prevails.

The words discendea, discernea, giacea, intendea, piacea, ravvolgea, reflettea, tenea, vedea and vincea, as well as many occurrences of other verbs at the imperfect are in a similar situation.

A very interesting case is

che non parëa s'era laico o cherco.	Inferno XVIII, 117

We accept Petrocchi's difficult choice to suggest an exceptional dieresis on "parëa" because there are no elements (rithmical, linguistical, nor according to Dante's usus scribendi) to privilege other solutions, equally exceptional:

che |non |pa|rea |s'e|ra |lä|i|co o |cher|co

che |non |pa|rea |s'e|ra |lai|co |o |cher|co

Lai|co (normally bisyllabic) could support an exceptional dieresis thanks to the Greek LAÏKÓS, Latin LÄICUS. While parëa can rely on the alternative frequent forms of the imperfect in -eva (pa|re|va).

C ANOMALOUS VERSES

There are just a dozen of cases in the Divine Comedy where the syllabification algorithm is raising a warning due to the absence of an accent on either the 4th or the 6th syllable. We have been glad to discover that all of them have been treated and discussed in the secular literary-critical tradition.

When an adverb in -mente is involved, we can enforce a secondary stress on the first component of the compound word (canìna-mente, miràbil-mente, glorïòsa-mente):

con tre gole caninamente latra	Inferno VI, 14
e vidila mirabilmente oscura.	Inferno XXI, 6
cotanto glorïosamente accolto.	Paradiso XI, 12

and a similar hypothesis (sustàn-zïàl from sustànza) can be advanced for

Ogne forma sustanzïal, che setta	Purgatorio XVIII, 49

The remaining verses are just considered as hendecasyllables with anomalous accentuation (2nd and 8th, or 3rd and 8th syllable):

mi pinser tra le sepulture a lui,	Inferno X, 38
parea che di quel bulicame uscisse	Inferno XII, 117
le lagrime, che col bollor diserra,	Inferno XII, 136
per lo furto che frodolente fece	Inferno XXV, 29
la vipera che Melanesi accampa,	Purgatorio VIII, 80
e 'Beati misericordes!' fue	Purgatorio XV, 38
e Cesare, per soggiogare Ilerda,	Purgatorio XVIII, 101