Metrics for the Deep Comedy project

Introduction

Italian metre and hendecasyllables

A formal definition of poetry is based on metrical quantities which do not appear in prosaic text. Although some experts suggest to discriminate between “poetic function” (the measurable attributes of a poetic work alone) and “true poetry” (encompassing also subjectively defined aesthetic factors), for the purposes of automatic generation, we will focus only on the former.

Unlike classic metres (ie. Greek and Latin) which are quantitative, Italian metres are accentative metres (ie. they are entirely defined by the accent positions inside the syllables of verses, or, more precisely, as the set of properties which can be derived by a sequence of tonic and atonic syllables).

Both accents and syllables considered in metre computations are rhythmic, not grammatical accents/syllables, meaning that there is not an exact correspondence with the way a poetic text is accented/syllabified with respect to a prosaic one[1]. For example, both the following verses have exactly 11 metric syllables, but 11 and 13 grammatical syllables, respectively:

- “Nel mezzo del cammin di nostra vita”
- “Esta selva selvaggia e aspra e forte”.

There are a set of phenomena (which can be considered as objectively legitimate requests made by the poet to the audience) which cause this discrepancy, the most common ones are:

- Synalepha
- Dialepha (for peculiar aspects related to dialepha in poesie toscane see[2])
- Syneresis
- Dieresis.

We will focus only on synalepha and dieresis, being the statistically most relevant and the easiest to detect, without performing a “constraint satisfaction” analysis on a poem (as it’s typically done by experts when dealing with a fully rigorous syllable count).

When words are spoken, the pauses between them have a duration which is not fixed and changes with context (eg. when a sentence is completed, pronunciation tends to emphasize a long pause, which is reflected by the use of punctuation in writing). When a word ends in a vowel and the next one begins with a vowel as well, a speaker tends to reduce the pause considerably (up to removing it completely when speaking very fast).

---

[1] F. Bausi, M. Martelli. La metrica italiana. Casa Editrice Le Lettere, 2nd edition, 1998, pages 11–13. isbn: 8871661362.

[2] F. Bausi, M. Martelli. La metrica italiana. Casa Editrice Le Lettere, 2nd edition, 1998, page 14. isbn: 8871661362.

Synalepha are a metric formalization of this of phenomenon. Since musicality is a requirement of poetry, synalepha are applied very often, to the point that they can be considered the “default” way of reading a verse intended by a poet.
In the previous example, three synalepha can be identified: “Esta selva selvaggia_e_aspra_e forte”, but (as only an expert can tell without counting syllables or considering cesure[3]) when declamating the verse, a pause between “e” and “aspra” needs to be added (ie. a dialepha nullifies the “hypothetical” synalepha).

When metric needs arise, the poet can ask a reader to vary the rhythm, not only by modulating pauses between words (ie. synalepha and dialepha), but also inside each word. As it will be clarified later, sequence of vowels in Italian can be pronounced either as a single sound (diphthongs/triphthongs) or as separate sounds (hiatus). When the pronunciation is changed from the intended meaning, the phenomena of dieresis (a diphthong is read splitting the sounds) and syneresis (a hiatus is joined pronouncing a single sound) arise. Since there is no way for the reader to infer such requirement (even knowing the number of syllables in the verse, there may be multiple interpretations of the poet’s will possible), a diacritic mark is placed over the vowels subject to these phenomena (eg. “sapïenza” means that it should be read as /sapiˈɛntsa/, instead of the usual /saˈpjɛntsa/[4], with, as it will be seen later, a different syllabification as well).

A terzina[5] is a strophe of three verses which can be coupled with any verse and rhyming scheme. A terza rima[6] is a terzina of hendecasyllables (although some poems written with a different metre are sometimes considered terza rima as well) rhyming according to the rima incatenata pattern, ie. ABA, where B will then rhyme with the first verse of the following terzina (as a result the last terzina in a canto will always be completed by a lone hendecasyllable[7]).

As already mentioned, Italian metres are not quantitative, therefore a specific metre is defined by the position of rhythmic accents inside it. All words, except clitics[8], possess their own grammatical accent, however, some words inside each verse are stressed more, in order to achieve rhythm. The rhythmic accents are the ones belonging to such words.

---

[3] "ceṡura in Vocabolario - Treccani." http://www.treccani.it/vocabolario/cesura/. Ultimo accesso: 9 giu. 2020.
[4] Other than the shift of the stress, also note the transition from the semivocalic sound /j/ to the vocalic sound /i/.
[5] The English translation is “tercet”, however we are refering to terzina/e, intending the specific characteristics of the Italian metre.
[6] "terza rima in "Enciclopedia dell'Italiano" - Treccani." http://www.treccani.it/enciclopedia/terza-rima_(Enciclopedia-dell'Italiano)/. Ultimo accesso: 9 giu. 2020.
[7] This will cause every rhyme to appear a multiple of three times, the number of verses to be divisible by two and many other patterns related to numerological symbolism arise.
[8] "clitici in "Enciclopedia dell'Italiano" - Treccani." http://www.treccani.it/enciclopedia/clitici_(Enciclopedia-dell'Italiano)/. Ultimo accesso: 9 giu. 2020.

Hendecasyllables[9] have evolved as a result of the merging of a quinario and a settenario, as such, they are defined as the metre having:

- A rhythmic accent either in the fourth or sixth syllable (from the quinario),
- A cesura separating the two “ancestral” verses,
- A rhythmic accent in the tenth syllable (from the settenario).

Since the tenth syllable’s accent is always the accent of the last word, it coincides with a grammatical accent and it can be considered itself a more informal (but still valid) definition for the metre.

In spite of being named after the number 11, it’s definition is clearly length-agnostic and, as such, it can have a variable number of syllables, from ten (endecasillabo tronco), up to sixteen (endecasillabo polisdrucciolo). The reason it has been given this name is the fact that most words have their accent on the penultimate syllable (parole piane) and as a consequence, if the penultimate syllable of a verse is the tenth, the verse has eleven syllables (endecasillabo piano).

For a gentle introduction to Italian metres, see[10].
For more technical insight, specific to the Divine Comedy, refer to[11].

Italian syllabification rules

A syllable is a group of phonemes which are emitted as a single sound. Any human language splits words into syllables in a process known as *syllabification*, this process imposes constraints on the combinations of phonemes allowed in a given language.

As a universal principle, a syllable must contain at least a *vocoid* and, optionally, antevocalic and/or postvocalic *contoids*[12]. A single syllable is phonologically a variation of intensity in the sound emitted by the larynx, with *peaks* corresponding to vocoids and *troughs* corresponding to contoids, syllabification can therefore be analyzed phonetically by isolating each trough-to-trough segment of the voice intensity[13]. Albeit simple and universal, this approach cannot be applied directly to written text, since the mapping between phonemes (sound units) and graphemes (text symbols) is not bijective, however in Italian vocoids always correspond to vowels (and therefore the first, and most important syllabification rule, can already be

[9] "endecasillabo in "Enciclopedia dell'Italiano" - Treccani." http://www.treccani.it/enciclopedia/endecasillabo_(Enciclopedia-dell'Italiano)/. Ultimo accesso: 9 giu. 2020.
[10] "La metrica italiana - Accademia Alfieri." http://www.accademia-alfieri.it/pagine/metrica.htm. Ultimo accesso: 8 giu. 2020.
[11] "endecasillabo in "Enciclopedia Dantesca" - Treccani." http://www.treccani.it/enciclopedia/endecasillabo_%28Enciclopedia-Dantesca%29/. Ultimo accesso: 8 giu. 2020.
[12] T. De Mauro. Linguistica elementare. Editori Laterza, 9th edition,1998, pages 35–37. isbn: 8842069779.
[13] . M. Federico Albano Leoni. Manuale di fonetica. Università / 357. Carocci editore, 3rd edition, 2002, pages 74–76. isbn: 9788843021277.

derived: a syllable must always contain at least a vowel) and the mapping presents very few exceptions.

For syllabification of a specific language, it's useful to use less general rules. Luckily, for the Italian language, these rules are well defined and only few cases of ambiguity arise when applied without phonetic informations (namely the position of the tonic accent inside words), moreover, heuristics exist for reducing these ambiguities further.

Before describing the syllabification rules for Italian, it's important to clarify some terms which will be used:

- Digram: sequence of two letters corresponding to a single sound, namely:
  - gl + i, gn + vowel,
  - sc + e/i,
  - ch + e/i, gh + e/i,
  - ci + a/o/u, gi + a/o/u (except when the i is accented, as in farmacìa).
- Trigram: sequence of three letters corresponding to a single sound, namely:
  - gli + a/o/u,
  - sci + a/o/u.
- Semiconsonant: vowel which is pronounced as if it were a consonant, can be only i/u in one of the following cases:
  - u + a/e/i/o
  - a/e/o/u + i,
  - a/e + u.
- Triphthong: sequence of three vowels pronounced with a single sound, namely:
  - iài, ièi,
  - uài, uòi, (eg. guài),
  - iuò (eg. aiuòla).
- Hiatus: sequence of two/three vowels pronounced separately, namely:
  - a + e/o (eg. maestro),
  - e + a/o,
  - o + a/e,
  - ì/ù + vowel(s) (eg. farmacìa, which is also an exception to the digram definition),
  - ri/bi/tri + vowel (eg. biella).

From the above terminology, two sources of ambiguity *arising when the accent is not marked* can be already be identified:

- ci/gi + vowel can be either a digram (and so will be pronounced with the same sound as the vowel) or one of the possible hiatus cases (i + vowel, pronounced as two separate sounds),
- i/u + vowel can be either a diphthong or a hiatus.

Heuristics able to solve these ambiguities with an high precision exist, even when the accent is not explicitly marked[14], however since hiatuses are statistically rare in current Italian (pronunciation of languages tends towards simplification over time, and diphthongs are easier to pronounce than hiatuses), in case of ambiguity, our algorithm will default to considering two consecutive vowels a diphthong.

The following rules are derived from [15] [16]:

- At the beginning of a word, a vowel or a diphthong form a syllable on their own (eg. **a**-mi-co, **au**-gu-ri, **e**-lo-qu-io),
- Diphthongs/triphthongs always belong to the same syllable (eg. a-
- **iuo**-la, u-ra-n**io**, ci-l**ie**-gia),
- Hiatuses are always split (eg. m**a-e**-stro, in-no-c**u-o**, in-vi-d**ï-o**-si, qu-**ï-e**- tar-mi),
- Digrams/trigrams are always together with the following vowel (eg.
- **ci**-lie-**gia**, **gno**-mo),
- A single consonant cannot be on it's own syllable (eg. **pa-lo**, a-**mi-co**),
- Double consonants (including cq) are always split (eg. a**c-q**ua, so**q-q**ua-dro, ri-du**r-r**e),
- Impure s (s + consonant) are always together with the following vowel (eg. **sce**l-le-ra-to, **sco**-**sta**r-si, **stra**p-pa-re),
- Other consecutive consonants are always split (eg. ac-cre-sci-me**n-t**o, a**n-t**ro-po-lo-gi-co), except when a mute consonant (b, c, d, g, p, t) is followed by a liquid one (l, r) (eg. pi-ro-**cla**-**sto**, ac-**cli**-ma-ta-to, an-**tro**-po-lo-gi-co, ac-**cre**-sci-men-to, s**tr**ap-pa-re), but not vice versa (cfr. a**r-t**o and **tr**op-po).

As stated in the previous section, synalepha and dieresis directly affect the syllabification process:

- If two words are joined by a synalepha, the last syllable from the first one and the first syllable from the second one are merged (even if they formed a hiatus under normal conditions, eg. "selva oscura" is syllabified as "sel-vaos-cu-ra", even if "a-o" always forms a hiatus)
- If a diphthong/triphthong contains a dieresis, it becomes a hiatus (eg. "sapïenza" is syllabified as "sa-pï-en-za", not "sa-pien-za").

[14] "iato in "Enciclopedia dell'Italiano" - Treccani." http://www.treccani.it/enciclopedia/iato_(Enciclopedia-dell'Italiano)/. Ultimo accesso: 7 giu. 2020.
[15] M. Sensini. Lo spazio linguistico, volume A (La riflessione sulla lingua). Arnoldo Mondadori Scuola, 10th edition, 2005, pages 15–17. isbn: 9788824724838.
[16] L. Serianni. Grammatica, sintassi, dubbi. Garzanti Editore, 2nd edition, 1997, pages 37–38. isbn: 8811504880.

Metrics

Syllable count

As discussed in the previous sections, an hendecasyllable doesn't have a count of exactly 11 syllables, being shorter if it ends with a parola tronca and longer if it ends with a parola sdrucciola (or polisdrucciola).
Since the accent's position can be inferred only heuristically (or by dictionary lookup), we considered the following two heuristics (useful also for rhyme detection):

- If a word ends with an accented vowel or is a monosyllabic word, it certainly is tronca,
- If a word ends with a consonant, it's probably tronca,
- Otherwise is piana (statistically the most significative case).

These heuristics perform statistically well (as it can be noted below, measuring the dataset's hendecasyllabicness), but there are notable cases like the following[17], which are all **valid** hendecasyllables with more than 11 syllables (even considering synalepha):

- "Sotto la penna, ovvero stalagmitificanomisi" (16, endecasillabo pentasdrucciolo),
- "Ottima è l'acqua; ma le piante abbeverinosene" (15, endecasillabo quadrisdrucciolo).

Rhyme detection

The tradeoff between precision and recall when detecting rhymes is extremely sensitive to tuning, due to the fact that some rhymes require a very short matching (eg. "f**u**"/"pi**ù**") while others a very long one (eg. "and**àndosene**"/"freg**àndosene**"[18]). In order to overcome this problem, we assign a matching score between two words, which is normalized with respect to the length of the substring which should match and then threshold that score in order to determine whether the two words rhyme.

Two words rhyme if there is an exact match from the tonic accent to the end (eg. "c**ane**"/"p**ane**"), under the assumption of knowing the accent position, match is trivial, however Italian grammar forces to mark the accent position with a diacritic only if it's located on the last letter.
Since accent estimation needs to be performed heuristically, a strict "rhyme"/"doesn't rhyme" outcome cannot be determined (if the accent position in two words is estimated before the real position, two rhyming words will be classified as false negative, conversely if the estimation is after the real position, two assonances will be classified as false positives).
To overcome this problem we assign a score for the queried words, which will be then thresholded to determine whether the words rhyme or not.

Given two words, there are four possible cases:

---

[17] "Endecasillabo - Wikipedia." https://it.wikipedia.org/wiki/Endecasillabo. Ultimo accesso: 8 giu. 2020.
[18] "BISDRUCCIOLA, ACCENTAZIONE in "La grammatica italiana"." http://www.treccani.it/enciclopedia/accentazione-bisdrucciola_%28La-grammatica-italiana%29/. Ultimo accesso: 6 giu. 2020.

- Both accents are known and on the same position from the end of the word: trivially match from the accent and onwards (if the match is perfect, the two words rhyme),
- Both accents are known and are on different positions: the two words trivially cannot rhyme,
- The accent is known only on one of the two words: mark the unknown accent at the same position of the known one and proceed as in the first case (this heuristic introduces some false positives, eg. “farmacìa”/”camicia” would incorrectly mark “camicìa” instead of “camìcia”, and the two words would wrongly be marked as rhymes)
- Neither accent is known: a fully heuristic estimation needs to be performed.

For the heuristic estimation we determine whether the word under examination is tronca or piana (we don’t have any heuristic for devising sdrucciole or polisdrucciole, since they are statistically infrequent, they are not correlated with lengthy words[19], and are rare as well at the end of a verse). Then we syllabify the word and consider only the last but one (if the word is piana) or last (if it’s tronca) syllable.

Since an accent can only be applied to a vowel (and since a syllable is always built around a vocoid core), we drop the consonants and look for a suitable vowel:

- If there is only one vowel, the choice is trivial,
- If there are multiple vowels (because the vocoid is a diphthong or a triphthong), we choose arbitrarily the last one, in order to give a more permissive match (we assume all diphthongs to be ascending[20]).

After having determined the length of the match, a matching score needs to be given. In order to achieve better empirical tunability (by setting the rhyming threshold) in terms of precision-recall tradeoff, a score more sophisticated than an hard 0/1 assignment is needed. Although many approaches can be taken, the most important property for the score is to be “normalized” with respect to the length of the match (so that the aforementioned “and**àndosene**”/”freg**àndosene**” and “f**u**”/”pi**ù**” will have similar scores). Another important property is to weight characters positionally (so that a mismatch “ànte”/”ènte” is less penalized than a mismatch “ànte”/”ànne”).

Our approach is to exponentially weight each character, starting from the end of the string and stopping to the first unmatched character (or the accent position if the match is perfect).

$$rhyme\ score\ =\ \frac{\sum_{i=0}^{first\ mismatch}\left(2^{minlength-i}\ if\ w1[minlength-i]=w2[minlength-i]\right)}{2^{minlength}}$$

Where minlength is the minimum distance from the accent to the end of the word for the two words (it can be different, eg. in “farmacìa”/”camìcia”, in which the distances are respectively 4 and 2 characters[21]). It can easily be shown that for a perfect match, this score is exactly 1

[19] eg. “precipitevolissimevolmènte”, the longest Italian word, is piana.
[20] "dittongo-ascendente: documenti, foto e citazioni nell ...." http://www.treccani.it/enciclopedia/tag/dittongo-ascendente/. Ultimo accesso: 8 giu. 2020.
[21] In this case we have a false positive, but this approach provides robustness with wrongly placed accents.

and that characters closer to the end of words are weighted more (guaranteeing a form of robustness even in the case the accent location was estimated to be farther than it should).

Local metrics

We analyze the text with a sliding window approach and extract each terzina conforming to the following structure:

```
verse
verse
verse

verse
```

For each valid terzina, we compute two metrics:

- Hendecasyllabicness: average compliance of each of the four verses to the hendecasyllable pattern,
- Rhymeness: considering the previous terzina as well, how many rhymes match the structure [xBx] BCB C divided by three (the transitivity of the second verse from the previous terzina matching the third verse of the current one is not checked).

Due to the problems highlighted in the previous sections, the two local metrics are configured by parameters:

- If the `synalepha` flag is set to `True`, syllables are split considering synalepha between words, before counting them
- The `permissive` flag changes the hendecasyllabicness score in the following way (where `target` is 11 for versi piani and 10 for versi tronchi):
  - $1.0 - \frac{|target - syl|}{target}$ if set to `True` and $|syl - target| > 1$
  - 1.0 if set to `True` and $|syl - target| \leq 1$
  - $1.0 - \frac{|target - syl|}{target}$ if set to `False`.
- The `rhyme_threshold` value determines when two words are considered to be rhyming (the default value of 1.0 is the strictest, decreasing it will cause assonances to be detected as rhymes).

Global metrics

We propose to use five global metrics to evaluate the performance of our network, these however cannot replace human-based judgement[22].

Number of putative terzine

Since each terzina has four lines (three verses and an empty line) and the last terzina has a stray verse, the number of hypothetical terzine an input with n lines can have is:

$$floor(\frac{n-1}{4})$$

[22] Unless, as stated at the beginning, we consider them as an evaluation for poetic function alone, instead of poetry. In that case, we believe they are strict and precise enough for the task.

Number of actual terzine

Every time the sliding window on the input detects a well formed terzina, a counter is increased.

Structuredness

$$\frac{number\ of\ actual\ terzine}{number\ of\ putative\ terzine}$$

Average hendecasyllabicness

$$\begin{cases} \frac{\sum_{\forall\, t \in actual\ ter} hendecasyllabicness(t)}{number\ of\ actual\ terzine} & \text{if } number\ of\ actual\ terzine > 0 \\ 0 & otherwise \end{cases}$$

Average rhymeness

$$\begin{cases} \frac{\sum_{\forall\, t \in actual\ ter} rhymeness(t)}{number\ of\ actual\ terzine} & \text{if } number\ of\ actual\ terzine > 0 \\ 0 & otherwise \end{cases}$$

For validation purposes, we evaluated our metrics over the first canto of Inferno, getting the following results (synalepha=True, permissive=False, rhyme_threshold=1.0):

```
Number of putative terzine: 45
Number of well formed terzine: 45
Average structuredness: 1.0
Average hendecasyllabicness: 0.9403535353535349
Average rhymeness: 1.0
```

Index