

# Deep Comedy

Lorenzo Mario Amorosa - 0000948133

Sami Mohammed Osman - 0000929408

## Introduction to the Divine Comedy

The Divine Comedy is an Italian masterpiece written by Dante in the XIV century. It is made of hendecasyllabic verses, all grouped in *terzina*, and it has a fixed rhyme scheme. Several previously written masterpieces, such as the *Iliad* and the *Odyssey*, have a strict and well known metric scheme as well.

This stylistic choice can be understood by thinking about that fixed and well thought out metrics facilitate the handing down of the work. In antiquity works were recited orally and hence a fixed rhyme scheme could help actors to remember phonetically words during the acting itself.

In addition to both rhymes and well thought out metric structure, images have a fundamental role in the Divine Comedy, in particular through figures of speech, such as metaphors and allegories [1]. As an example, we now take a look at *I canto* of *Inferno*:

- Dante initially gets lost into *selva oscura*, metaphor of *sin*
- then he meets 3 *beasts*:
  - *the lion* allegory of *sensuality*
  - *the lion* allegory of *pride*
  - *the wolf* allegory of *greed*
- then he meets *Virgilio*, allegory of *human reasonableness*, who proposes himself as Dante's travel guide

Each one of *100 canti* of the Divine Comedy is plenty of information which hints to images, in particular it is plenty of *lexemes*, which are atomic concepts, which meticulously describe reality and which are semantically highly detailed.

It has to be recognized that metaphors and allegories have a fundamental and central role in the masterpiece, indeed they bind indissolubly together images and words. The information is surely denser in words or in a *terzina*, but it is explicitly clearer in the associated image.

It must be mentioned that Dante's topics are often really controversial, e.g. *the sin*, *human protagonism*, *politics*, *Virgilio*, and metaphors and allegories are widely used to describe them.

## Deep Comedy - ideally

A good principle that could lead to the implementation of a neural network capable of generating a *canto* of the Divine Comedy must take into account the previously described facts. The neural network should ideally learn to associate words in a given *terzina* and its rhymes to the relative images.

At the state of the art it is scarcely feasible, but now we will illustrate our attempts to face the problem.

## Deep Comedy - for real

We initially started to experiment with really simple models, such as plain RNN, LSTM and GRU, which in other scenarios led to impressively good results[2], but the performances were not acceptable. As a consequence, we decided to implement more complex models, such as Transformers[3], which performed really well.

The zip file we provide contains the code needed for the implementation of the Transformer capable of generating a canto, and it is organized as follows:

DeepComedy.zip

```
divine_comedy.txt      # The raw input data
Final_DeepComedy.ipynb # The notebook containing the code to run
                        # the Transformer
metrics                # The folder with the code to evaluate the
                        # performance
model.h5               # The weights of the model
subword_text_encoder.py # Utility module for subword tokenization
results                # The folder with the results of the
                        # evaluation metrics varying network and
                        # dataset hyperparameters
```

To execute the code it is enough to open the Python notebook in a Jupyter environment. In the following sections of the report we will illustrate the main content of the notebook, more details can be found in the notebook directly.

## Data Preprocessing

First of all, we preprocessed the raw data to give a more organized structure to the text, in order to make it easier for the Neural Network to understand the tokenized text. In particular, we insert some special tags in the original Divine Comedy which are:

- =SOCA= for the start of a cantica
- =EOCA= for the end of a cantica
- =SOC= for the start of a canto
- =EOC= for the end of a canto
- =TER= between each couple of tercets

Here is an example of a sample of text before and after the tagging:

<p>INFERNO - Canto I</p> <p>Nel mezzo del cammin di nostra vita mi ritrovai per una selva oscura, ché la diritta via era smarrita.</p> <p>Ahi quanto a dir qual era è cosa dura esta selva selvaggia e aspra e forte che nel pensier rinova la paura!</p>	<p>=SOCA= =SOC= Nel mezzo del cammin di nostra vita mi ritrovai per una selva oscura, ché la diritta via era smarrita. =TER= Ahi quanto a dir qual era è cosa dura esta selva selvaggia e aspra e forte che nel pensier rinova la paura! =TER=</p>
---	--

## Text Tokenization

We choose a subword level tokenization which is responsible to take care not only of alphanumeric symbols but also of punctuation symbols, new-line symbols, white spaces and special tags, all separate tokens that are independently learnt and generated by the Neural Networks.

The subword level tokenization is preferred because it performs better than the char level or word level tokenization for the hendecasyllabicness metric. Indeed subword tokenization and hyphenation often split words in the same places, hence it is easier to keep the syllabic scheme. The most frequent tokens are extracted from the Divine Comedy.

Here is an example of the subword tokenization of a sample of text:

<p>=SOCA= =SOC= Nel mezzo del cammin di nostra vita mi ritrovai per una selva oscura, ché la diritta via era smarrita. =TER=</p>	<p>=SOCA= =SOC= Nel mez zo del cam min di nos tra vit a mi rit rov ai per una sel va osc ura , ché la dir itt a via era sma rri ta . =TER=</p>
--	--

## Dataset Hyperparameters

We tried to use a *Sequence To Sequence* model, usually used for Neural Machine Translation, to generate a Dante's canto. Even though these kind of models are generally used to translate from one natural language to another, we wanted to apply the same idea to the Divine Comedy, trying to "translate" a sample of verses into the entire following verses. Accordingly, the dataset is built to have a fixed number of verses as input and a target set of verses as output.

The hyperparameters in the construction of the dataset are:

- the window size to extract the input strings, or rather the *sequence length*
- the number of tokens to skip from one window to the following one, or rather the *step length*
- the train/validation splitting percentage
- the batch size

Our experiments resulted in the choice of *sequence length* equals to 3 and *step length* equals to 1. In this way both valuable information about the rhyming scheme is not wasted and all the dataset is fully exploited. Indeed, from the previous three verses (included the one having as a single token the =TER= tag) the network should be able to understand both the structure and the rhyming scheme, namely it should be able to understand whether the next verse should contain a =TER= tag or not, and if not how to properly end the verse so that it rhymes when necessary.

### Optimizer and Loss Function

The Transformer model was compiled with a custom Adam optimizer, whose specifications for a better optimizer were given in the original paper [3]. Instead, as loss function we used the Sparse Categorical Crossentropy.

### Model

The Transformer [3] is a state-of-the-art architecture proposed by Vaswani et al. in 2017 in order to overcome the problems encountered by RNNs for Neural Machine Translation tasks, namely the poor capacities in long-term memory and attention, and the high computational cost. Additional reasons why we decided to try this architecture are due to its great ability to generalize in different Natural Language Processing tasks as well as its remarkable training speed.

The Transformer consists of an encoder and a decoder, both made of a variable number of layers which affects the overall complexity of the model. It is not a recurrent network but, instead, it is a standard feed-forward one. As a consequence, the input tokens of the decoder and the encoder must be processed not only by a traditional embedding layer but also by a so-called positional encoding, so that the information about the order of the tokens in the sequence is not forgotten even without the presence of recurrent edges.

In the end, these processed inputs are passed through the series of layers, which are made up of two different sub-modules: a multi-head attention and a subsequent feed-forward block. In particular, the decoder is made of two multi-head attention layers, one responsible for processing the decoded input, the other responsible for processing the merged vectors of features extracted both from the encoder input and the decoder input. Everything is eventually passed through a dense layer, with a softmax activation on top, to obtain the output probabilities of each token.

### Model and Training Hyperparameters

The hyperparameters of the model and of the training phase are:

- the number of layers for the encoder and the decoder
- the number of heads for the multi-head attention sub-module
- the dimension of all the sub-layers in the model, as well as the embedding layers, known as  $d_{model}$
- the inner feed-forward layers dimension, known as  $d_{ff}$
- the dropout rate
- the number of training epochs

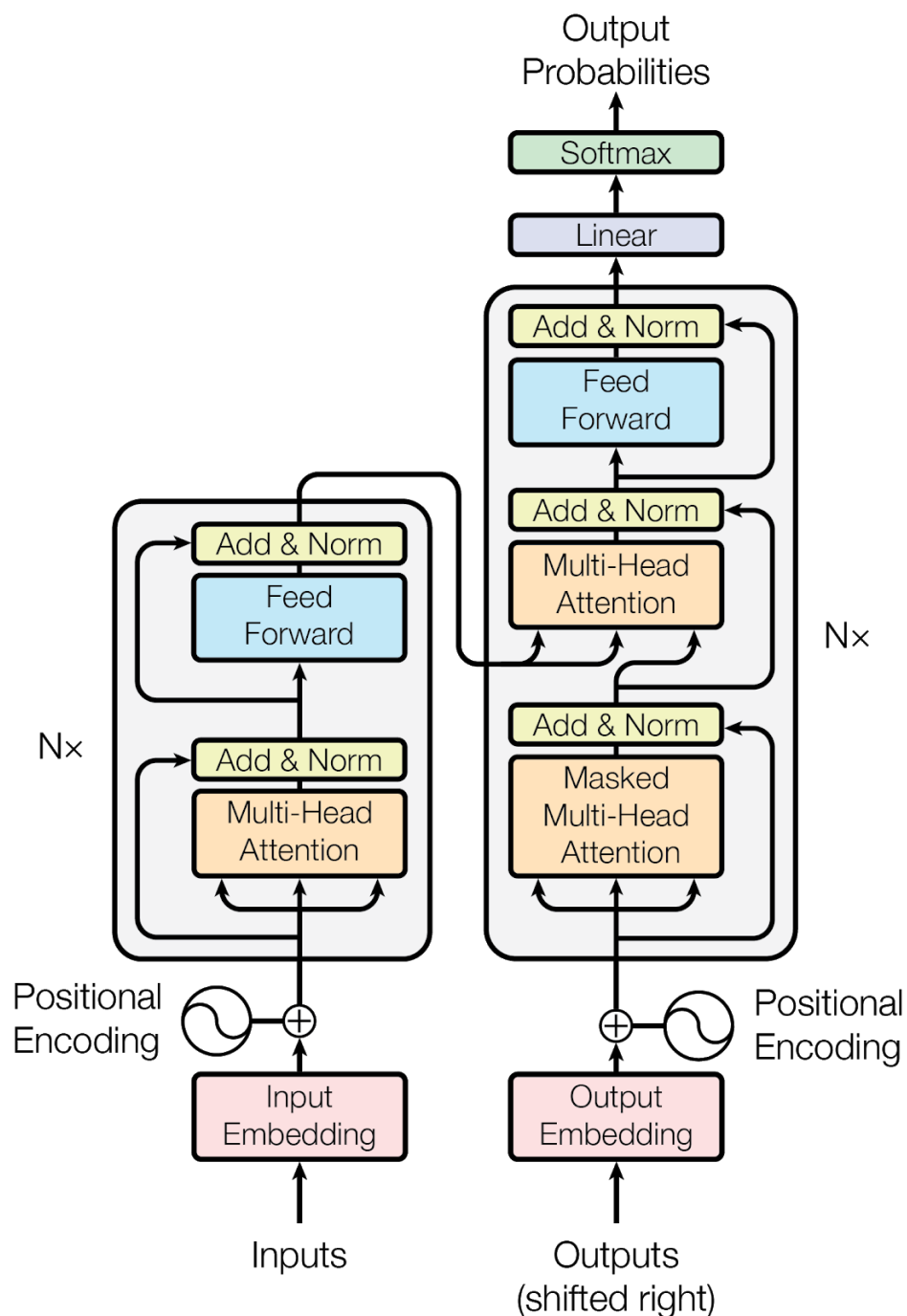


Fig.1. Transformer architecture

## Results

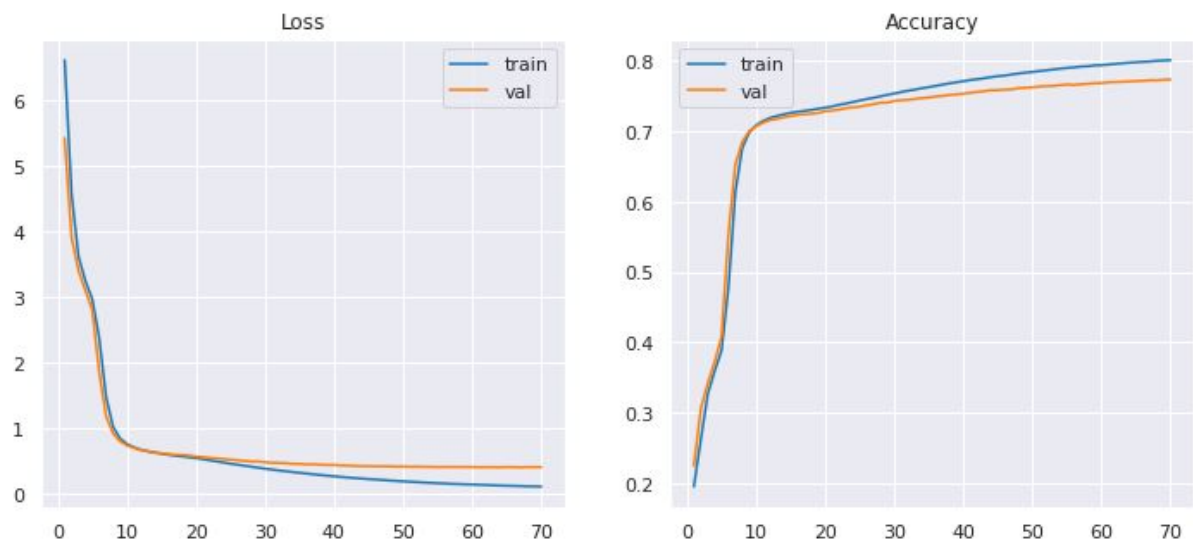
We experimented with several combinations of previously described hyperparameters, whose results can be found in the zip file. In the end, accordingly to the evaluation metrics, the best set of them is:

```
# Dataset Hyperparameters
seq_length = 3
```

```
step_length = 1
batch_size = 64
train_val_split = 0.7

# Model and Training Hyperparameters
num_layers = 3
num_heads = 4
d_model = 256
dff = 512
dropout = 0.2
epochs = 70
```

Here's a graphical representation of the improvement of the model, with respect both to the loss and the accuracy, across the epochs, using the best hyperparameters:



*Fig. 2. Loss and accuracy graphs*

Here the attention plot of two heads on some verses is shown to understand how the model behaves. The whiter is the point in the graph, the higher is the importance given to the input token. Input tokens are on the horizontal axis, output ones are on the vertical axis

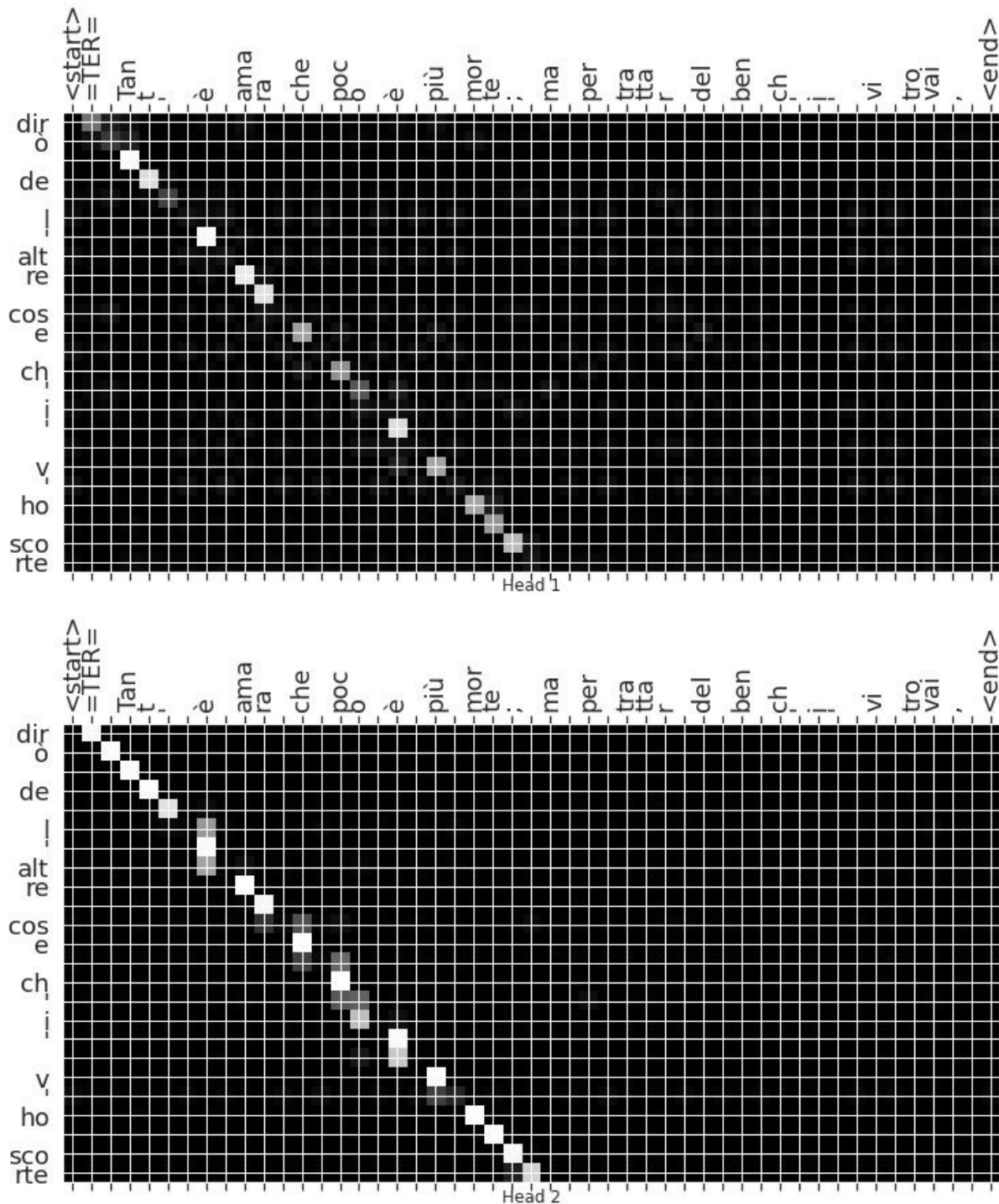


Fig. 3. Attention plot

Here the results of the evaluation metrics applied on the generated canto are shown:

Hendecasyllabiness:	0.904,
Plagiarism:	0.959,
Repetitiveness:	0.998,
Rhymeness:	0.896,
Text Structuredness:	0.996,

Putative Tercets: 62,  
Well-Formed Tercets: 62.

In the generation of a canto, it is possible to allow a certain degree of randomness using a *temperature\_factor*. The next token is chosen among a subset of those having a probability which is at least  $1 / \text{temperature\_factor}$  with respect to the maximal one. As a consequence, a higher *temperature\_factor* leads to a more explorative generation, while a lower *temperature\_factor* leads to a more conservative one (in particular, with *temperature\_factor* of 1 the generation is completely deterministic).

In the end, the entire canto, generated with a *temperature\_factor* of 1.0, is proposed:

per me' chii' chini,  
ma 'l tuo sommo bene è tanto,  
si disse: "Maestro, or mi puoi il mio paterno;

e cerchio il suggel cor santo stretto  
de la virtù che fa conviene al segno,  
e al tratto mio ben l'innocenti al petto.

In quel che piace a me l'hanno in ira  
fosse in quel che ci ha fatto in ira,  
secondo che Dio convien che più spira.

Se venir sù l'allor vi disira",  
disse 'l maestro mio senza parte accende,  
"quel bene ha mestier di là si tira".

E io l'accorgor; né lo ritene,  
ho io d'i pianeti che far l'esser legge,  
né ferma qui pena che non m'accende.

E 'l savio come me stesse cade  
de la mente lor di quei che 'l ciel chiaro  
tutte quante a la quattro strade.

Con più oltre era ancor non senza giuro,  
così per esser lor passaver remi  
poscia che nel viso suo ver giuro.

Ma quel digiunto, del nostro mondo  
sen va in dietro, e 'l viso girone  
del tuo maestro e io disir mi secondo;

e quel che più lieve si può fare,  
s'io non era ancor di là dal volto,  
come fa di noi per la grazia non si chiude.

Ma quella vedta tutto quanto aspetto  
che colpa negasse quel che t'assotta,  
prezioso mia mente mia benedetta,



e la vostra natura e voi siete  
onesto voi natura, e asti e voi lieta  
la voi che la persona e la spira.

Tu se' sì pietà perché l'lieta cade  
che 'l loco ove s'infiamma e ciò guarda,  
che dal cor sovra nostro amor vade.

Ver è ch'altra fiata quanto non ti sponda  
la carne che s'infiamma e sparse,  
e qualite ad una altra imagina.

Noi ci vidi due col sol di cose  
eravamo imagine come l'ultima ripa,  
la donna mia così ferita imposte.

Quella circolazion che sì dura  
facevan lui far di sé sarebbe,  
come solo il sol mio inganna.

E come quel cui venimmo al girale  
si volge intorno, come l'occhio intorno  
qual è più grande e più non si sale".

Quanto più torcer di là è corpo orno;  
e questo passar per lo gran distinto,  
poscia che 'l cerchio di quel di addorno.

E come l'infima quanto era quinto  
di ciò che fosse il suo fattore tempio,  
e non per lo lume del vento di quanto;

ma come a guarda e vidi costui,  
per l'orribil sabbion di là possente,  
vid' io verso noi là sù ver' esso.

Però disse: "Tu vedrai gente ride  
se non fosti a come tu degna,  
sì che 'l mortal divina voglia,

e come 'l quale in sù non si convegna  
de l'eterno valore, e qual io vidi  
quand' io mi traggon d'una parte degna.

Lo duca mio intende, e tutti altri giri  
di ciò che dal convenio corso è procede  
la bocca e con sì dolcemente e spiri.

Mentre che andar per l'occhio sì fiere,  
che pria lor veder prima e lor maggiori,  
e libente di Dio non fosse sole.

Vedi che la mia mente se tu di spli,  
dicendor: "Signor mio, e tu non torni

ch'i' non potrei or chi ti disnodi".

"O mio padre mio, "sanza servo torni",  
rispuos' io, "e io cominciai ed esso,  
"tra 'l gran non conosciuto li pieni.

"O navicao del piacere o pavento  
che suoli il capo mio in su la aperta",  
disse 'l maestro, "io li dissi e 'l sento".

Ond' elli a lui: "Se tu non ti raccerta  
lo duca mio a cui la meta al tratto  
de la nostra occulta mente mia scorta;

e tal girando il mio cor de l'agutto  
di sovra sé un punto da cammino,  
quand' io mi volsi al suo dir tutto.

Non vidi sùbito lor visto e speraro,  
che fosse in una nave gridò: "Perché fine  
le tue parole il nostro e nel chiaro".

E io a l'un duca mio di lei accese;  
per ch'io mi parsi e l'altro parlare,  
lo sol va un poco a l'affisse e come 'l prende.

E come l'altre facea ch'i' mi parve  
dimando per quella luce che ne l'un l'aspetto  
che per udir come fa che 'l dolce.

Sì come de l'altra vita diletto  
per veder se tu così lo scoglio,  
come tu fosti, in altro ben diletto.

E come quel può nessun si raccoggio  
di quel cui trestizzo fiammeggiar m'accorgo  
l'ardor de li occhi al nostra regno;

e anche l'occhio le gingambe il viso  
non vide un, ma per colui che 'l cerchio  
coruscia tra 'l capo e con l'altro grosso.

Lo maggior de l'aguglie racceso  
si misero dal corpo de l'ali,  
che non si misura al suo valor mi rimfiso.

Lo buon maestro innanzi a me sali,  
non so 'l pur innanzi che non fuor giunto,  
se non fosse ch'i' non l'altro bassi.

Quando fuor de la grazia ch'i' ho scovra  
dinanzi al senso ingrato, ch'ogne parte  
che l'arte del cor m'ha raggio in ira.

E come l'uom che fosse l'arte e l'arte,  
sì che già l'occhio m'avea tutto l'inferno  
del ben cesso giù de le sue ir mi rime.

Se 'l duca mio e de l'eterno roso  
di quel fummo in su dal nostro stanca  
e basciavansi per tutto e 'l fosso sogco.

Quando col viso e con la cui s'accorga  
lo suo tormento a le nostre perfidanza,  
si misea com' avea già cotanta.

Così com' io da quella parola paranza  
a l'acqua che la gente segue la fronte  
che non avesse il primo grave caranza.

Sì come l'altra matera ci diserrante,  
per altre rocca per la bocca senza tese,  
sì che 'l duro ti discerno sante.

Questi avea membro tutte l'uom cortese  
e ristetro al mondo dal suo concetto,  
e d'un mento che l'altro facea quel gele.

E quel che più facea l'uom più del cento,  
e qual io dissi: "Maestro, ed esso  
che l'altro è d'uno e l'altro intender pieto.

E se non fosse che 'l cielo è messo  
che par per me l'uno in su l'altro strinse  
per lo segno che m'ha fatto per lo specchio.

Qual è colui che più strinse pinse  
col viso in su per lo scoglio riguardo  
ne l'un punto fiso non li occhi s'accorse;

tal qual io a l'occhio per l'impgiocondo  
che quel cagion che non si raccolse  
l'alta colpa che non si nascondo al mendo.

Quivi certo io nel viso sù, voler si scese,  
per avea virtù con tanto disse:  
"O tu che non può esser non ti fiese".

E io a cui volo: "Maestro, assai morse  
vostra ragion non s'aspetta e s'aspetta,  
se li occhi pur come a me satisfaccse;

e chi potrem dentro a' terra diletta,  
sì che 'l ben tra 'l terzo s'aspetto,  
e quando i prieghi grande è più procetta?

Per ch'ogne mia vista de l'occhio aspetto,  
s'è quel dubbio che m'è tanto avviva,

e io mi dissi: "Maestro, or mi diletto

ne la mia scorta quanto più è più luna;  
e quella scender qua sù nel mondo,  
con la prima sua la voglia prora.

Tanto ti parver m'ebbe a te secondo  
discese qual si sanpuò ciascuno, e vedervi,  
"Le cose nel ben tutto quanto.

Oh potenza dal signor quarto miran divini  
non per bene ad opere che divenne  
a l'eterno aspetto che s'aspetta.

Queste ascoltar chi due chiavi conte:  
ché pria non lo veder qual era allor mi presso  
come quello amor ch'i' non mi conte.

Né per lo 'nferno a tanto basso,  
come prima ch'el fu detto la prima corda,  
tu veder di cui venir ci trassi.

Ma ché tu chi fosti era la goda  
più volto io più calar con altro mondo,  
ben è lunga scala sua ben bene".

Sì come l'un tacer non m'ha riso;  
e però che 'l presente non si sponda,  
quando la bocca e non divina aschio.

Mentre ch'io forma mi distava: "Ella;  
l'ombra di carro, e di sopra l'ombra  
di quel reverenza di sùbito d'una.

Noi discendemmo il Virgilio ed el s'inferna,  
che di là discese ne l'un tace,  
quando la dolce del loco m'ha cerchia;

e s'io appreso questo con quella pace  
che, s'appresso in paradiso, si cerna  
di quel cui si adde l'ultimo cane.

Così sì ch'io mirò questo sospesa  
di contrario mi si chiama in grido,  
come per cerchio che si convenia.

E io: "Maestro, or mio, or m'al profondo  
mio cominciar mostra il viso e cerchio,

## **Bibliography**

[1] Jacomuzzi, Dughera and Ioli. "La Divina Commedia", 2012.

[2] Andrej Karpathy. "The unreasonable effectiveness of recurrent neural networks". In: Andrej Karpathy blog 21 (2015), p. 23.

[3] Ashish Vaswani et al. "Attention is all you need". In: Advances in neural information processing systems. 2017, pp. 5998–6008.