"Il Mago della Ghigliottina" @ Ghigliottin-AI: When Linguistics meets Artificial Intelligence

Federico Sangati[1,2], Antonio Pascucci[1], and Johanna Monti[1]
[1]L'Orientale University of Naples - UNIOR NLP Research Group, Italy
[2]Okinawa Institute of Science and Technology Graduate University, Japan
federico.sangati@gmail.com, {apascucci,jmonti}@unior.it

Abstract

English. This paper describes *Il mago della Ghigliottina*, a bot which took part in the *Ghigliottin-AI* task of the Evalita 2020 evaluation campaign. The aim is to build a system able to solve the TV game "La Ghigliottina". Our system has already participated in the Evalita 2018 task *NLP4FUN*. Compared to that occasion, it improved its accuracy from 61% to 68.6%.

Italiano. *Questo contributo descrive Il mago della Ghigliottina, un bot che ha partecipato a Ghigliottin-AI, uno dei task di Evalita 2020. Scopo del task è mettere in piedi un sistema automatico capace di risolvere il gioco televisivo "La Ghigliottina". Il nostro sistema ha già partecipato all'edizione del 2018 di Evalita al task NLP4FUN. Rispetto all'edizione del 2018 di NLP4FUN, l'accuratezza è salita dal 61% al 68.6%.*

1 Introduction

In this paper we describe *Il mago della ghigliottina* (Sangati et al., 2020), a bot which participated in *Ghigliottin-AI*, one of the Evalita 2020 tasks (Basile et al., 2020a). Evalita[1] (Basile et al., 2020b) is an initiative of AILC (Associazione Italiana di Linguistica Computazionale) and is a periodic evaluation campaign of Natural Language Processing (NLP) and speech tools for the Italian language, which takes place every two years in conjunction with CLiC-IT[2], the Italian Conference on Computational Linguistics. *Ghigliottin-AI* takes its cue from the Evalita 2018 *NLP4FUN* (Basile et al., 2018) task. Participants are asked to build an artificial player able to solve "La Ghigliottina", the final game of the popular Italian TV quiz show "L'Eredità". The game involves a single player, who is given a set of five words (clues), unrelated one to each other, but related with a sixth word that represents the solution to the game. Our system took already part in the 2018 Evalita task *NLP4FUN* as *UNIOR4NLP* (Sangati et al., 2018). *Il mago della Ghigliottina* is identical to *UNIOR4NLP*, being based on the same principles and methodologies: analyzing real game instances we found out that in most cases clues and solution are connected because they form a Multiword Expression (MWE). A MWE can be defined as a sequence of words that presents some characteristic behaviour (at the lexical, syntactic, semantic, pragmatic or statistical level) and whose interpretation crosses the boundaries between words (Sag et al., 2002). MWEs are lexical items which convey a single meaning different from the meanings of the constituents of the MWE, such as in the idiomatic expression *kick the bucket* where the simple addition of the meanings of *kick* and *bucket* does not convey the meaning of *to die*. We have decided to participate as *Il mago della ghigliottina* instead of *UNIOR4NLP* because after participating in the *NLP4FUN* task in 2018 we developed three different versions of the solver *Il mago della ghigliottina* available as i) a Telegram Bot (@Unior4NLPbot)[3], ii) a Twitter bot (@UNIOR4NLP) and finally iii) an Amazon Alexa skill (Mago della Ghigliottina). This paper is organized as follows: in Section 2 we present related work and in Section 3 we provide an overview of the task. In Section 4 we describe our system. Results are shown in Section 5 while in Section 6 we focus on the error anaysis. Conclusions are in Section 7 along with future work.

[1]http://www.evalita.it

[2]http://clic2020.ilc.cnr.it/it/home

[3]A short video showing how the bot works is available at https://youtu.be/3fggGlJaSII

2 Related work

Languages have always been a source of inspiration to create games. As the years passed, the possibility to rely on large linguistic resources and artificial intelligence has allowed scholars to build systems able to solve games, which represent an interesting playground to test the results of research (Yannakakis and Togelius, 2018). When we think about linguistics and artificial intelligence it is almost obvious to think to the IBM Watson system, which successfully challenged human champions of Jeopardy!^{TM}, a game where contestants are presented with clues in the form of answers, and must phrase their responses in the form of a question (Ferrucci et al., 2013). Another interesting example is represented by solvers of Italian crosswords (Ernandes et al., 2008; Littman et al., 2002). The scientific community periodically organizes i) shared tasks to evaluate Natural Language Processing (NLP) applications in the solution of linguistic games (*Ghigliottin-AI* is an example) and ii) workshops focused on games and gamification for NLP tasks. The *Games and NLP* (Lukin, 2020) workshop, for instance, was organized this year in the context of the LREC 2020 conference. Fourteen teams presented their research in occasion of this workshop, and according to the submitted papers, we can state that the research moves in two directions: i) the exploitation of NLP techniques to solve linguistic games on the basis of semantic relations between words and ii) the development of *Games With A Purpose (GWAPs)* in order to crowdsource linguistic data from engaged players.

TV games, such as "*Wheel of Fortune*", "*Who wants to be a Millionaire?*" and, indeed, "*La Ghigliottina*" represent an interesting test bench for linguistic knowledge-based systems. (Molino et al., 2015) exploit question answering techniques to build an artificial player for *Who wants to be a Millionaire?*. With regard to our specific case study, other systems were built to solve "*La Ghigliottina*". OTTHO (Semeraro et al., 2009; Basile et al., 2016), the first artificial player of "*La Ghigliottina*", is a system based on i) web resources (e.g. Wikipedia) in order to build a lexicon and a knowledge repository and ii) a knowledge base modeling represented by an association matrix which stores the degree of correlation between any two terms in the lexicon. Word correlations are detected by connecting i) lemmas to the terms in its dictionary definition, pair of words occurring in a proverb, movie or song title, and ii) pair of similar words by exploiting Vector Space Models (Salton et al., 1975). During the *NLP4FUN* Task in 2018 two systems took part in the competition: our system (which is presented in Section 4) and (Squadrone, 2018), that proposed an algorithm based on two steps: i) for each clue of a game, a list of relevant keywords is retrieved from linguistic corpora, so that each clue is associated with keywords representing the concepts having a relation with that clue. Then, words at the intersection of the retrieved sets are considered as candidate solutions; ii) another knowledge source made of proverbs, book and movie titles, word definitions, is exploited to count co-occurrences of clues and candidate solutions. A further system developed to solve "*La Ghigliottina*" game is Robospierre (Cirillo et al., 2019), a system which relies on MWEs automatically extracted through a lexicalized association rules algorithm, on a list of proverbs and on some lists of titles.

3 The *Ghigliottin-AI* task

Ghigliottin-AI is one of the Evalita 2020 tasks. The aim of Evalita (which in 2020 reached its seventh edition) is to promote the development of language and speech technologies for Italian, providing a shared framework where different systems and approaches can be evaluated in a consistent manner. *Ghigliottin-AI* participants are asked to build an artificial player able to solve "La Ghigliottina", the final game of the Italian TV show "L'Eredità". Given a set of five words (clues) the player has to find the solution to the game which is a sixth word related with each one of the five clues. The five clues are unrelated one to each other. For example, given the set of clues *conoscere* (*to know*), *grado* (*degree*), *modello* (*model*), *ideale* (*ideal*) and *divina* (*divine*) the solution is *perfezione* (*perfection*) because: *conoscere alla perfezione* (*to perfectly know*), *grado di perfezione* (*degree of perfection*), *modello di perfezione* (*model of perfection*), *ideale di perfezione* (*ideal of perfection*) and *perfezione divina* (*divine perfection*). In order to train participants' systems, the task organizers provided a set of 300 games with their five clues and their solution in a JSON format. This training set is taken from the last editions of the TV game. The systems have been then evaluated using an API

based methodology, namely the Remote Evaluation Server (RES) *Ghigliottiniamo*[4] which currently enables both humans and artificial systems (bots) to submit solutions to the TV game in real-time. The test set consists in 350 games instances, provided by *Ghigliottiniamo* at random intervals of time as a request with a single game challenge to registered systems. The RES allowed systems to reply with a single solution to the game. Similar to the original TV game, where players have 60 seconds to provide the solution, the RES discards solutions received after 60 seconds from the submitted challenge. The same happened in evaluating systems participating in *Ghigliottin-AI*.

4 System description

This section describes *Il mago della Ghigliottina*, which, as already mentioned, is the system submitted in 2018 without any changes. The system is based on the analysis of real game instances: in most cases clues and solution are connected because they form a MWE. A further observation is that clues are always nouns, verbs or adjectives, while solutions are nouns or adjectives. On this basis, we have detected six patterns that identify MWEs connecting clue/solution pairs:

A B pattern: *diario segreto* ('diary secret' → secret diary), *brutta caduta* ('ugly fall' → bad fall), *permesso premio* ('permit price' → good behaviour license), *dare gas* ('give gas' → accelerate).

A det B pattern: *dare il permesso* ('give the permit' → authorize).

A prep B pattern: *colpo di coda* ('flick of tail' → last ditch effort).

A conj B pattern: *stima e affetto* (esteem and affection).

A prepart B or **A prep det B** pattern: e.g. *virtù dei forti*, part of the famous Italian proverb *La calma è la virtù dei forti* (patience is the virtue of the strong).

A+B pattern: compounds such as *radio* + *attività* = *radioattività* (radio + activity = radioactivity).

The system is based on a number of freely available corpora:

Paisà : 225 M words corpus automatically annotated (Lyding et al., 2014).

itWaC : 1.5 B words corpus automatically annotated (Baroni et al., 2009)

Wiki-IT-Titles : Wikipedia-IT titles downloaded via WikiExtractor[5].

Proverbs : 1955 proverbs from Wikiquote[6] and 371 from an online collection[7].

In addition, we have developed the following lexical resources:

DeMauro-Ext : words extracted from "Il Nuovo vocabolario di base della lingua italiana"(De Mauro, 2016b), extended with morphological variations obtained by changing last vowel of the word and checking if the resulting word has frequency ≥ 1000 in *Paisà*.

DeMauro-MWEs : MWEs extracted from the "De Mauro online dictionary" (De Mauro, 2016a) composed of 30,633 entries.

More technical details about our system are available in (Sangati et al., 2018), submitted for the *NLP4FUN* task.

5 Results

In this section, we discuss results and we also compare the performances achieved by our system in *Ghigliottina-AI* with those achieved in the Evalita *NLP4FUN* task. Compared to our participation in *NLP4FUN*, when our system proved to be the best performing one (Basile et al., 2018), the accuracy has increased from 61.0% to 68.6%. This is probably due to the fact that while the 2020 edition only used games from the TV game, in the 2018 edition 39 out of the 105 games in the test set were taken from the board game. This supports what already reported in (Sangati et al., 2018), that is, the board game edition presents different types of word-association as compared to the TV game. The Table 1 provides the performances of our system in both editions of the task.

[4] https://quiztime.net

[5] http://attardi.github.io/wikiextractor. Last accessed on the 1st October 2018

[6] https://it.wikiquote.org/wiki/Proverbi_italiani. Downloaded on the 24th April 2018

[7] http://web.tiscali.it/proverbiitaliani. Downloaded on the 24th April 2018

Task	Correct	Total	Accuracy
Ghigliottin-AI (2020)	240	350	68.6%
NLP4FUN (2018)	64	105	61.0%

Table 1: Results on the *Ghigliottin-AI* and the *NLP4FUN* TEST sets. In the column "Total" we show the number of game instances in the test set. Accuracy is computed as the number of correct games divided the total.

6 Error analysis

In the attempt of providing the correct solution to the 350 game instances that compose the *Ghigliottin-AI* test set, 110 errors have been made, which represent 32.4% of the whole test set. In this section we discuss the errors, trying to analyze and justify them. In particular, we try to detect the motivation behind errors, in order to categorize them. The following list presents examples of different categories of errors we detected.

6.1 High correlation between clue(s) and our solution.

One or more clues have a high correlation with the wrong solution provided by the system.

A clues: *fare* (*to do*), *saldo* (two different meanings *sale* and *balance*), *interessato* (*interested*), *grande* (several meanings, such as *big* and *great*) and *attenzione* (*attention*). Our system provided the solution *shopping* (the same in English) instead of the right one *richiesta* (*request*). In this case the system didn't disambiguate correctly the meaning of *saldo* (*richiesta di saldo*, namely *balance request*). The system chose the solution *shopping* instead of *richiesta* due to the high correlation between *shopping* and *saldo* (*sale*). One possible explanation is that *shopping* and *saldo* almost always occur in the same sentence. For this reason the solution *shopping* achieved a higher weight compared to that of other solutions;

B clues: *brutto* (*ugly*), *fare* (*to do*), *morto* (*dead*), *cavaliere* (*kinght*) and *diavolo* (*devil*). The solution is *paura* (*fear*), while our system provided the solution *povero* (*poor*). Considering that our system is also trained with a list of proverbs, in this case the error is due to the high correlation between *povero* and *diavolo* (*povero diavolo*) is a famous way of saying;

C clues: *perdere* (*to lose*), *amicizia* (*friendship*), *bottiglia* (*bottle*), *acqua* (*water*) and *quattro* (*four*). The right solution is *segno* (*sign*), but our system provided the solution *bicchiere* (*glass*) due to the high correlation with *bottiglia* and *acqua*.

6.2 Right kind of reasoning, wrong solution.

Wrong solutions such as singular instead of plural (and vice-versa), or trivial mistakes in the face of a right kind of reasoning.

D clues: *questione* (*question*), *indagine* (*investigation*), *disegno* (*design*), *pagamento* (*payment*) and *lavorare* (*to work*). Instead of *metodo* (*method*), our system provided its plural *metodi*;

E clues: *copertina* (*cover*), *dimensione* (*dimension*), *persona* (*person*), *seno* (*sinus*) and *età* (*age*). The solution is *terza* (*third*), but our system provided the wrong solution *quarta* (*fourth*) which has correlation with all the five clues.

6.3 Clue(s) and solution are synonyms.

The solution provided by our system is a synonym of one or more clue(s).

F clues: *essere* (*to be*), *prezzo* (*price*), *fermo* (*stop*), *capitale* (*capital*) and *regolare* (*regular*). The solution is *partenza* (*departure*), but our system provided the solution *fisso* (*fixed*) which can be intended as a synonym of *fermo*.

6.4 Unclear solutions.

This subsection discusses some strange solutions provided by the system. The solutions that our system detects are listed from the best to the worst one. Then, it chooses the best one. Thanks to a debug function it is possible to analyze the solutions provided by the system in order to understand what is their correlation with one or more clues. The examples provided below concern solutions apparently strange which we analyzed thanks to this function.

G clues: *vecchio* (*old*), *cavallo* (*horse*), *end* (the same in English), *soda* (the same in English) and *conquista* (*conquest*). The solution is *west* (the same in English). Our system provided the solution *polenta* (the same in English), which is a dish as well as the surname of a famous Italian commander lived in the 13th century (Guido da Polenta), also knows as "il Vecchio" (the Elder);

H clues: *gioco* (*game*), *trovare* (*to find*), *fuori* (*out*), *dollaro* (*dollar*) and *quadrato* (*square*). The solurion is *area* (the same in English), but our system provided the solution *straccio* (*shred*), due to the high correlation with *trovare* because of the way of saying *non trovare uno straccio di prova* (*do not have a shred of evidence*);

I clues: *erba* (*grass*), *sangue* (*blood*), *indagine* (*investigation*), *prova* (*evidence*) and *miss* (*Miss*). The solution is *campione* (*champion*), but our system provided the solution *pazienza* (*patience*), because of the high correlation with *erba*: *Erba pazienza* (*Patience Dock*) is the common name for the Rumex patientia plant.

Debugging the system also allows us to observe if the right solution is in the list of best solutions provided by the system and how it is ranked. Statistics based on the 110 errors recorded during the test phase are reported in Table 2, where "best of 5" means: best solutions detected when there is correlation between each one of the solutions and all the five clues. The same reasoning applies to "best of 4" and "best of 3".

Correct solution is	Occurrences
the 2nd best solution	22
in the Best of 5 list	30
in the Best of 4 list	13
in the Best of 3 list	6
not in the list	61

Table 2: Correct solutions in our system list when error solutions are provided as best solution

As we can see, in 22 cases the correct solution is the second best solution detected by our system. In 61 cases the correct one is not in the whole list of possible solutions detected by the system.

6.5 Part-of-speech errors.

We also noticed that some errors are due to the selection of solutions with a wrong part-of-speech (POS). In Table 3 we report the occurrences of POS errors.
In particular, in 26 cases the system selected an adjective as solution instead of a noun, for example:

J - clues: *scrivere* (*to write*), *rosso* (*red*), *luce* (*light*), *colori* (*colors*) and *inchiesta* (*inquiry*). The solution is *film* (*movie*), namely a Noun. In this case while our system provided the solution *giallo* (*yellow*) (an Adjective). We can also note that the error solution has been provided because two of the five clues (*rosso* and *colori*) are related to the same conceptual group of *giallo*, namely colors.

Error POS	Correct POS	Occurrences
Noun	Noun	80
Adjective	Noun	26
Noun	Adjective	2
Verb	Noun	2
Noun	Adjective	-

Table 3: Occurrences of error POS provided by our system instead of correct POS

7 Conclusions and future work

In this paper we described *Il mago della ghigliottina*, a system which took part in the Evalita 2020 *Ghigliottin-AI* task. Our system achieved an accuracy of 0.6857, with 240 correct solutions given on a test set composed of 350 game instances. As already mentioned, our system is the same system which took part in the Evalita 2018 *NLP4FUN* task and is designed on a key observation: clues are connected to the solution because they form a multiword expression (MWE). In order to build our system, we collected linguistic and lexical resources described in Section 4. Since future work will focus on improving the performances of the system, a special focus has been dedicated to error analysis. Section 6, in fact, presents different categories of errors we detected (with examples and clarification of errors) as well as statistics about correct solutions presence in our system list of solutions.

Acknowledgements

This research has been partly supported by the PON Ricerca e Innovazione 2014/20 fund. Authorship contribution is as follows: Federico Sangati is author of Sections 4 and 5, Antonio Pascucci is author of Sections 3 and 6, Johanna Monti is author of Sections 1 and 2, Abstract, Conclusions and future work are in common.

References

Marco Baroni, Silvia Bernardini, Adriano Ferraresi, and Eros Zanchetta. 2009. The wacky wide web: a collection of very large linguistically processed web-crawled corpora. *Lang. Resources and Evaluation*, 43(3):209–226.

Pierpaolo Basile, Marco de Gemmis, Pasquale Lops, and Giovanni Semeraro. 2016. Solving a complex language game by using knowledge-based word associations discovery. *IEEE Transactions on Computational Intelligence and AI in Games*, 8(1):13–26.

Pierpaolo Basile, Marco de Gemmis, Lucia Siciliani, and Giovanni Semeraro. 2018. Overview of the evalita 2018 solving language games (nlp4fun) task. In Tommaso Caselli, Nicole Novielli, Viviana Patti, and Paolo Rosso, editors, *Proceedings of the 6th evaluation campaign of Natural Language Processing and Speech tools for Italian (EVALITA'18)*. CEUR.org, Turin, Italy.

Pierpaolo Basile, Marco Lovetere, Johanna Monti, Antonio Pascucci, Federico Sangati, and Lucia Siciliani. 2020a. Ghigliottin-ai@evalita2020 evaluating artificial playersfor the language game "la ghigliottina". In *Proceedings of Seventh Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2020)*.

Valerio Basile, Danilo Croce, Maria Di Maro, and Lucia C. Passaro. 2020b. Evalita 2020: Overview of the 7th evaluation campaign of natural language processing and speech tools for italian. In Valerio Basile, Danilo Croce, Maria Di Maro, and Lucia C. Passaro, editors, *Proceedings of Seventh Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2020)*. CEUR.org, Online.

Nicola Cirillo, Chiara Pericolo, and Pasquale Tufano. 2019. Robospierre, an artificial intelligence to solve" la ghigliottina". In *CLiC-it*.

Tullio De Mauro. 2016a. Il Nuovo De Mauro (Online). https://dizionario.internazionale.it. Last accessed on the 1st October 2018.

Tullio De Mauro. 2016b. Il Nuovo vocabolario di base della lingua italiana (pdf version). https://www.internazionale.it/opinione/tullio-de-mauro/2016/12/23/il-nuovo-vocabolario-di-base-della-lingua-italiana. Last accessed on the 1st October 2018.

Marco Ernandes, Giovanni Angelini, and Marco Gori. 2008. A web-based agent challenges human experts on crosswords. *AI Magazine*, 29(1):77–77.

David A. Ferrucci, Anthony Levas, Sugato Bagchi, David Gondek, and Erik T. Mueller. 2013. Watson: Beyond jeopardy! *Artif. Intell.*, 199:93–105.

Michael L Littman, Greg A Keim, and Noam Shazeer. 2002. A probabilistic approach to solving crossword puzzles. *Artificial Intelligence*, 134(1-2):23–55.

Stephanie M. Lukin, editor. 2020. *Workshop on Games and Natural Language Processing*. European Language Resources Association, Marseille, France.

Verena Lyding, Egon Stemle, Claudia Borghetti, Marco Brunello, Sara Castagnoli, Felice Dell'Orletta, Henrik Dittmann, Alessandro Lenci, and Vito Pirrelli. 2014. The PAISÀ corpus of italian web texts. In *Proceedings of the 9th Web as Corpus Workshop (WaC-9)*, pages 36–43. Association for Computational Linguistics, Gothenburg, Sweden.

Piero Molino, Pasquale Lops, Giovanni Semeraro, Marco de Gemmis, and Pierpaolo Basile. 2015. Playing with knowledge: A virtual player for "who wants to be a millionaire?" that leverages question answering techniques. *Artificial Intelligence*, 222:157–181.

Ivan A Sag, Timothy Baldwin, Francis Bond, Ann Copestake, and Dan Flickinger. 2002. Multiword expressions: A pain in the neck for nlp. In *International Conference on Intelligent Text Processing and Computational Linguistics*, pages 1–15. Springer.

Gerard Salton, Anita Wong, and Chung-Shu Yang. 1975. A vector space model for automatic indexing. *Commun. ACM*, 18(11):613–620.

Federico Sangati, Antonio Pascucci, and Johanna Monti. 2018. Exploiting multiword expressions to solve "la ghigliottina". In *Sixth Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2018)*, volume 2263, pages 258–263. Accademia University Press.

Federico Sangati, Antonio Pascucci, and Johanna Monti. 2020. The challenge of the tv game la ghigliottina to nlp. In *Workshop on Games and Natural Language Processing*, pages 34–38.

Giovanni Semeraro, Pasquale Lops, Pierpaolo Basile, and Marco De Gemmis. 2009. On the tip of my thought: Playing the guillotine game. In *Proceedings of the 21st International Jont Conference on Artifical Intelligence*, IJCAI'09, pages 1543–1548. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

Luca Squadrone. 2018. Computer challenges guillotine: how an artificial player can solve a complex language tv game with web data analysis. *EVALITA Evaluation of NLP and Speech Tools for Italian*, 12:262.

Georgios N Yannakakis and Julian Togelius. 2018. *Artificial Intelligence and Games*. Springer.