

Exploring Lexical-Semantic Knowledge in the Generation of Novel Riddles in Portuguese

Hugo Gonalo Oliveira

CISUC, DEI

University of Coimbra, Portugal

hroliv@dei.uc.pt

Ricardo Rodrigues

CISUC

Polytechnic Institute of Coimbra, Portugal

rmanuel@dei.uc.pt

Abstract

We describe an effort towards the automatic generation of novel riddles in Portuguese, ultimately with humour value. Riddle generation fits in the common architecture of a NLG system and may follow different models, described here, all based on features of a concept, acquired from a lexical-semantic knowledge base. Generated riddles were manually assessed by humans, who rated them as fairly interpretable, surprising, and novel, even if with low humour potential.

1 Introduction

To act naturally, computers should be able to entertain, e.g., with the creation of wordplay or humour. This paper is about the automatic generation of novel riddles, which would ideally be apt for humorous contexts. Riddles are a kind of linguistic puzzle, present in most cultures and languages. They can be posed as a question, followed by a short hiatus — allowing for the audience to think —, and finally an answer that works as a punchline. Their generation is closely related to the topic of Computational Humour (Ritchie et al., 2006), which studies the utilisation of humour by computers, with value for Natural Language Understanding — specifically, developing models for automatically recognising and understanding humour (Mihalcea and Strapparava, 2006; Yang et al., 2015) —, and for Natural Language Generation (NLG) — developing systems that produce verbal humour (see section 2), which is our case.

Both riddles and humour have been generated by others, following different approaches, but often in English. Our work is inspired by the previous and represents an effort towards the development of a riddle generator in Portuguese. We exploit available lexical resources for this language, mainly a semantic knowledge base, and

implement six models for generating novel riddles, hopefully funny, due to the introduction of potentially incongruent word-play, forcing the re-interpretation of known concepts or the creation of new ones. This is a follow-up, as we have previously proposed the six generation models in a single-page paper (Gonalo Oliveira and Rodrigues, 2018) and analysed the results of different word sources, relations, presentation modes, and automatic scoring, for two of those models (Gonalo Oliveira and Rodrigues, 2018b).

We frame our current results as *piadas secas* (roughly, dry jokes) which, in Portugal, are a kind of joke that may be presented as a riddle, with a question and a short not-so-funny, obvious or nonsensical answer, taking advantage of the anti-climax to make people laugh. In 2017, this kind of jokes seemed to have had a comeback, as they were used in several television shows and YouTube videos, and re-compiled in websites or edited books (Pinto et al., 2017). This is why the system is baptised as SECO (dry).

The remainder of the paper starts with a brief overview on related work, covering the generation of riddles and verbal humour. The steps for the generation of riddles with the help of a knowledge base are then described, with some examples. Before concluding, we present the results of human validation of generated riddles, illustrated with additional examples. According to human subjects, riddles are accessible, surprising and novel, but, on average, not so funny.

2 Related Work

Riddles have long been a research topic. Georges and Dundes (1963) define them as “a traditional verbal expression which contains one or more descriptive elements, a pair of which may be in opposition; the referent of the elements is to be guessed.” Palma and Weiner (1992) confirm that lexical ambiguity, arising, for instance, from pol-

ysemy or homophony, is paramount to their creation, alongside the association of words to *ad-hoc* categories, according to their multiple meanings.

Riddle generation by computer programs has been addressed in the 1990s, with the seminal work of [Binsted and Ritchie \(1994\)](#), who developed JAPE, a system that generates punning riddles based on a syntactic and semantic lexicon; a set of schemata for combining two words based on their lexical or phonetic relationship; and a set of templates that render riddles as text. An example of a riddle would be “*What do you call a murderer that has fibre? A cereal killer.*” STANDUP ([Manurung et al., 2008](#)) adopts a similar approach, but is more cautious on the words used, restricted at various levels for a better output and suitability for the intended audiences.

TheRiddlerBot ([Guerrero et al., 2015](#)) generates riddles about famous characters. After selecting a well-known name from a knowledge base: associated features are retrieved; analogous characters, with common features, are identified; a textual template is selected for rendering the riddle, based on some of the features or the analogy; the riddle is posted in Twitter; and aliases are retrieved from Wikipedia, so that users may answer with the name of the selected character or one of its aliases. An example of a riddle for the Joker would be: “*Tell me the name of a person that is the Morpheus of The Dark Knight Rises, is criminal, playful yet cruel, has been seen wearing a purple topcoat.*”

Riddles have also been generated from word associations ([Galvan et al., 2016](#)). Given a concept: its possible categories are obtained from a creative thesaurus; associated modifiers are also retrieved and one is randomly selected; new categories, to which the selected modifier is associated to, are retrieved; a final category is composed by combining the selected modifier with one of the new categories; a concept of the final category is used to fill a text template. An example of a riddle for the sun would be “*What is as hot as soup?*” Unlike JAPE and STANDUP, this system and TheRiddlerBot do not tackle the humour aspect specifically.

The ConceptNet semantic network was exploited for the generation of verbal humour, including riddles ([Labutov and Lipson, 2012](#)). For this purpose, different, but overlapping, paths between the same two concepts are aligned with a surface template that maximises inter-path incongruity. For question-answer riddles, the question

mentions two concepts from different paths, but same domain, while the concept in the answer is in one of the paths, but from a different domain. An example of a riddle is “*Why is the computer in hospital? Because the computer has virus.*”

Besides riddles, there is work on the generation of other kinds of verbal humour, including funny acronyms ([Stock and Strapparava, 2006](#)), or short messages ([Valitutti et al., 2016](#)). Both explore lexical replacement for potentiating humour. Replacement words are constrained by the original form (same initial letter or similar sound) and possibly other humour-specific features (e.g., taboo).

All of the previous works generate riddles or humour based on some theory and knowledge resources, essential for their goal. They share a number of similarities with ours, but they all produce text in English. A few exceptions generate humour in other languages (e.g., [Sjöbergh and Araki \(2007\)](#)), but none generates riddles in Portuguese. In this language, however, there is work on the automatic generation of Internet humour ([Gonçalo Oliveira et al., 2016](#)), based on an image macro and a line of text.

Humour occurs when there is a break of conventionality in language. Understanding it is a sign of fluency ([Tagnin, 2005](#)), which explains the interest of computational linguistics on this topic. Humour is generally based on four linguistic phenomena, at the written and oral levels ([Tagnin, 2005](#)), namely: homonymy — words with the same spelling and sound, e.g., ‘band’, musical group or ring —, homophony — words with the same sound, but different spellings, e.g., cent and scent —, polysemy — words with the same spelling and sound, but multiple related meanings, e.g., wood, timber or forest —, and paronymy — words with similar spellings or sounds, e.g., collision and collusion. As such, its success depends highly on how proficient the audience is in the language used.

According to [Attardo \(2008\)](#), humorous plots can rely on one of the following: a punch line (typical jokes); a meta-narrative disruption, introducing familiar humorous references; or an otherwise normal situation, where some elements may cause laughter by the shortcomings of the agents.

3 Riddle Generation Approach

SECO explores lexical resources in Portuguese for generating word and feature combinations. Those can be rendered as novel riddles, to be used for

human entertainment purposes, and can ultimately have some humour value, and thus be seen as punning riddles. If we see them as jokes, following Attardo (2008)’s classification, SECO falls under the first type, as it presents a humorous plot (question) with a punchline (answer).

As in JAPE (Binsted and Ritchie, 1994), riddle generation may follow different implemented models, all resorting to a (lexical-semantic) knowledge base (KB), in order to acquire features of given lexicalised concepts. A parallelism can roughly be made between our approach for generating riddles and the common architecture of a NLG system (Reiter and Dale, 2000), as it encompasses the following steps:

- **Model Instantiation & Feature Acquisition** (roughly, Content Determination): sets the model and exploits the KB for a combination of initial concept and related features;
- **Riddle Creation** (roughly, Microplanning): selects the appropriate words for denoting different types of feature and sets how the selected combination is going to be presented as text, i.e., as a definition or as a question-answering pair;
- **Rendering** (Surface Realisation): renders the combination as text, after some adaptations that make it more natural.

This section describes the previous steps in more detail and ends with some examples.

3.1 Model Instantiation & Features

All generation models implemented start with an initial concept with two detachable parts (c_1 , c_2), either lexicalised as a compound (e.g., *human rights*) or a single word that, based on its orthography, may be divided in two (e.g., *knowledge=know+ledge*). Each part is considered individually and features are retrieved from the KB, some involving the first part of the concept (in set F_1), others the second part (F_2). Features are represented as triples of the kind *a relatedTo b*, where *a* and *b* are words and *relatedTo* is the name of a semantic relation between meanings of *a* and *b* (e.g., *animal* hypernymOf *dog*). As such, every feature in $f_1 \in F_1$ and $f_2 \in F_2$ will consist of a relation involving, respectively, c_1 and c_2 .

- $F_1 : \forall(x, relatedTo, y) \in F_1$
 $\rightarrow (x = c_1 \wedge y = fw_1) \vee (x = fw_1 \wedge y = c_1)$

- $F_2 : \forall(x, relatedTo, y) \in F_2$
 $\rightarrow (x = c_2 \wedge y = fw_2) \vee (x = fw_2 \wedge y = c_2)$

After this, features $f_1 \in F_1$ are paired with features $f_2 \in F_2$. The result is a set of combinations of the initial concept with pairs $\{fw_1, fw_2\}$. Figure 1 illustrates this step, with d_1 and d_2 representing textual descriptions of the features that connect c_1 to fw_1 and c_2 to fw_2 , respectively.

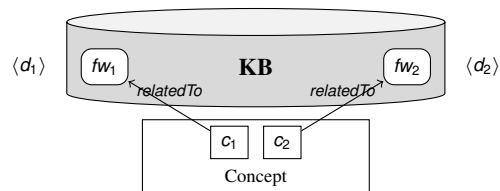


Figure 1: Feature acquisition

3.1.1 Lexical Resources

Two lexical resources are used by all the riddle generation models and are thus essential to this work. A **lexical-semantic knowledge base (KB)** with relation instances that occur in at least three out of ten semantic networks for Portuguese, including dictionaries, wordnets and Concept-Net (see Gonçalves Oliveira (2018) for additional details), that has a total of 45,510 instances, covering a rich set of relation types (see section 3.2), and is used by all the models.

A **morphology lexicon** (Ranchhod et al., 1999) with more than 900,000 Portuguese word forms, their part-of-speech and other grammatical information. It is used by the models to handle inflections and can also be used as a source of words (e.g., in models based on a single word).

3.1.2 Implemented Models

Inspired both by the schemata of JAPE (Binsted and Ritchie, 1994) and by the kind of jokes that Portuguese children use to make, we implemented six riddle generation models. A key difference is that they produce riddles in Portuguese and are thus constrained by the available lexical resources in this language. All models instantiate the generic model of Figure 1, but follow a different intuition, reflected in their input and, for the last two, on the features explored.

Reinterpretation of compounds (RC): given a known noun+adjective compound ($c_1 + c_2$), features are acquired for each of its words individually, and used to (re-)define it. Our intuition is that the meaning of the compounds is more than just

the sum of the meanings of both of their words, which may result in unexpected associations, possibly perceived as incongruent, and thus humour-prone. The input for this model was a list of 180 Portuguese noun+adjective compounds (Ramisch et al., 2016), with instances such as *água doce* (fresh water), *mau-humor* (bad mood), or *primeira mão* (first hand/leg).

New compounds (NC): explores the idea that humour may result from new words with familiar sounds (homophony, paronymy). It generates new compounds from all pairs of valid words with an edit distance of 1 letter to the original compounds, and then works as RC. With the same list as RC, instances like *amido oculto* (occult starch), *véu aberto* (open veil) or *primeiro pano* (first cloth) are obtained, respectively from *amigo oculto* (occult friend), *céu aberto* (open sky) and *primeiro plano* (first plan).

Reinterpretation of words (RW): instead of compounds, this model is based on single words, interpreted as a blend of two. This often leads to an unexpected meaning, attributed to the word, again perceived as incongruent, thus increasing the humour potential. Target words are in the lexicon and have the form $w_1 w_2$, where w_1 and w_2 are character sequences that are also in the lexicon. Instances of this kind include *malabar* (juggle) — interpreted as *mala+bar* (suitcase+bar) —, *centralidade* (centrality) — interpreted as *central+idade* (central+age) —, or *restolho* (stubble) — interpreted as *resto+olho* (rest+eye).

New blends (NB): analogous to NC, but in respect to RW instead of RC. Before trying to reinterpret words, a small change is made in their orthography, according to a short handcrafted list of possible character sequence replacements that do not change the sound of the word excessively. Those are sequences like $\{on, un, \tilde{a}o\}$, $\{i, e\}$ or $\{r, l\}$, which can be replaced in lexicon words and result in instances such as *fundasom* — obtained from *fundação* (foundation) and split into *funda* (catapult) and *som* (sound) —, *bombesta* — from *bombista* (bomber), split into *bom* (good) and *besta* (beast) —, or *calavela* — from *caravela* (caravel), split into *cala* (shut) and *vela* (sail). Each part of the instance (C_1 and C_2) must be in the lexicon, and thus be valid, but their blend ($w_1 w_2$) is not necessarily an existing word. The result is a new concept with a sound that resembles a known one, interpreted as the blend of

two other (often unrelated) concepts.

Partial antonyms (PA): assumes that the orthography of some words starts or ends with the antonym of another and that novel antonyms may result from changing the start/end of those words with its antonym. Instances of this kind include *pormenor* (detail) — where *menor* (smaller) is antonym of *maior* (bigger), resulting in novel antonym *pormaior* —, or *diante* (in front of) — where *dia* (day) is antonym of *noite* (night), resulting in antonym *noitente*. To avoid atypical words, some considerations on how syllables are formed and their limits are made. For instance, with a direct replacement, the antonyms of *sair*, *assalto*, *salvo* and *negativo* would be, respectively, *saan-dar*, *assbaixo*, *snegro* and *negpassivo*. With those considerations, they become *sandar*, *asbaixo*, *sal-negro* and *nepassivo*.

Antonymy Blend (AB): revisits the idea that, due to their orthography, some words can be interpreted as a blend of other two. Yet, it is focused on those where both parts (C_1 , C_2) have an antonym. This is used in the generation of novel antonyms, such as *odiar-atingir* (hate-reach) — for *amarfalhar* (to crumple), interpreted as *amar+falhar* (to love+to fail) —, *sombra-ilegal* (illegal-shadow), for *solicito* (solicitous), interpreted as *sol+lícito* (sun+lawful) —, or *mau-mau* (bad-bad) — for *bombom* (candy), interpreted as *bom+bom* (good+good).

3.2 Riddle Creation

Riddles have to be conveyed in natural language in a way that the relation between the concept and the features is understood. Depending on the types of feature, the text to clarify the relation is set, before selecting how the riddle is to be presented.

3.2.1 Feature Description

In the KB, relation instances connect words according to one of their meanings, but different meanings of the same word are not explicitly identified. Though unintentionally, this enables the selection of less expected meanings because the chance of selecting exactly the same meaning a word has in a compound is low, especially when the global meaning is not the sum of the parts (C_1 , C_2) meaning.

As the relations used are present in at least three other resources, they are generally well-known and consensual. If this constraint is dropped, more features can be extracted, though less immedi-

ate. The risk of using incorrect features is also higher, because most available semantic networks for Portuguese are created (semi-)automatically, with minimal or no curation.

From the available relations, we identified a subset of types that could be used as features. Some of those types are listed in Table 1, together with their frequency in the KB, and the text of their description, in Portuguese, to be used in the riddles, followed by a rough English translation. As only synonymy and antonymy are symmetrical relations, some types have different descriptions, depending on the position of the concept word (c_i) and feature word (fw_i) in the relation. There are also relations for which no text is necessary besides the feature word itself. Finally, if fw_i is a noun, it will be preceded by an indefinite article (*um*, for masculine, *uma* for feminine).

#	Relation	arg ₁	arg ₂	Description (d_i)
7,538	adj-synonymOf	fw_i / c_i	c_i / fw_i	<i>o que é fw_i</i> (what is fw_i)
353	adj-antonymOf	fw_i / c_i	c_i / fw_i	<i>o que não é/está fw_i</i> (what is not fw_i)
157	v-antonymOf	fw_i / c_i	c_i / fw_i	<i>o que não fw_i</i> (what does not fw_i)
4,035	hypernymOf	fw_i	c_i	fw_i
590	adj-saidAbout-n	c_i	fw_i	fw_i
100	n-partOf-adj	fw_i	c_i	<i>o que é fw_i</i> fw_i
58	n-partOf-n	c_i	fw_i	<i>uma parte de fw_i</i> (a part of fw_i)
		fw_i	c_i	<i>o que tem fw_i</i> (what has fw_i)
		fw_i	c_i	<i>o que tem fw_i</i>
110	v-purposeOf-n	c_i	fw_i	<i>a finalidade de fw_i</i>
		fw_i	c_i	<i>o que serve para fw_i</i>
1,572	v-causes-n	c_i	fw_i	fw_i
		fw_i	c_i	<i>o efeito de fw_i</i> (the effect of fw_i)

Table 1: Feature types and textual description

3.2.2 Presentation Mode Selection

Each combination of an initial concept and related features can be presented in different (hard-coded) ways, such as those in Table 2. Those include a definition (**DEF**); a question with two features for which the initial concept is the answer (**FC**); a question with the initial concept for which the features are in the answer (**CF**); or questions that ask for the opposite of the initial concept / pair of features, answered with a pair of features / initial concept (**OP1**, **OP2**).

3.3 Rendering

The riddle is finally rendered as a text with the concept and the features description. Yet to make text more natural, additional adaptations have to be made, depending on the presentation mode and

ID	Template
DEF	$\langle concept \rangle : d_1 f_1 e d_2 f_2$
FC	Que resulta do cruzamento entre $d_1 f_1$ e $d_2 f_2$? $\langle concept \rangle$. (What do you get when you cross ... ?)
CF	O que significa $\langle concept \rangle$? $d_1 f_1$ e $d_2 f_2$. (What does ... mean?)
OP1	Qual é o contrário de $\langle concept \rangle$? $f_1 f_2$. (What is the opposite of ... ?)
OP2	Qual é o contrário de $f_1 f_2$? $\langle concept \rangle$. (What is the opposite of ... ?)

Table 2: Presentation templates for riddles

on the kind of features, namely:

- A1 Noun features should appear before other features.
- A2 If one feature is a noun and the other an adjective, the description of the adjective is removed, because it is clear enough that the adjective is modifying the noun.
- A3 If one feature is a noun and the other is an adjective, the adjective must have the same gender as the noun and is inflected according to the lexicon.
- A4 If both features are nouns, *que é* (what is) is added before the second, and both can be interpreted as ‘co-hypernyms.’
- A5 If both features are adjectives, the description of the second is removed.
- A6 If one feature is an adjective but can also be used as a noun, it is used as a noun.
- A7 When necessary, the conjunction *e* (and) is added between the feature descriptions.

Adaptations A1 to A5 apply to the presentation modes DEF, CF and OP1; A6 is applied to FC; and A7 is applied to all.

3.4 Dissected Examples

To illustrate the previous steps, we present some examples of generated riddles. Consider, for example, the concept *direitos humanos* (human rights), in the compounds list, for which the features acquired from the KB include:

direito synonymOf *liso* (right, flat)
direito synonymOf *plano* (right, plane)
direito antonymOf *torto* (right, bent)
humano saidAbout *homem* (human, man)

From those, the riddles in Table 3 could be the result of the RC model. The instantiation of the generic model that results in the fifth riddle is depicted in Figure 2.

Despite different possible ways for presenting the riddles, a default one was set for each model, namely: FC for RC, except when one of the features is an antonym, with OP2 being used instead; CF for NC, RW, NB; OP1 for PA and AB. Table 4 has examples of riddles generated by each model, covering every rendering adaptation, and including a rough English translation.

Riddle	Pres.	Adapt.
<i>direitos humanos: um homem liso.</i> (human rights: a flat man)	DEF	A1, A2
<i>direitos humanos: um homem plano.</i>	DEF	A1, A2
<i>direitos humanos: um homem que não é/está torto.</i>	DEF	A1, A2
<i>Que resulta do cruzamento entre o que é liso e um homem? direitos humanos.</i>	FC	A7
<i>Que resulta do cruzamento entre um plano e um homem? direitos humanos.</i> (What do you get when you cross a plane and a man? human rights.)	FC	A6
<i>Qual é o contrário de homem torto? direitos humanos.</i>	OP2	A1, A2
<i>O que significa direitos humanos? um homem liso.</i>	CF	A1, A2
<i>O que significa direitos humanos? um homem plano.</i>	CF	A1, A2
<i>O que significa direitos humanos? um homem que não é/está torto.</i> (What does human rights mean? a man that is not bent.)	CF	A1

Table 3: Riddles generated for the concept *direitos humanos* (human rights)

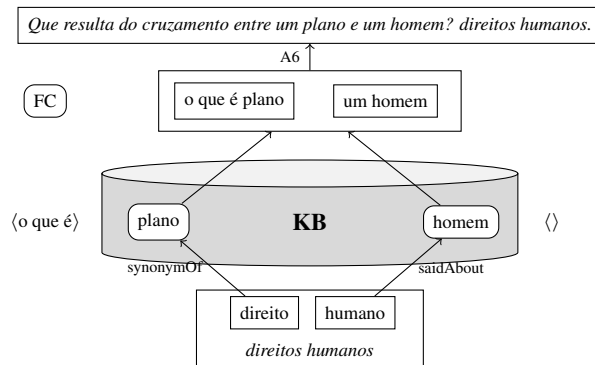


Figure 2: Instantiation of the RC model with the expression *direitos humanos*

4 Validation

The assessment of the riddles includes highly subjective aspects, namely humour. Also given that the riddles were created for human consumption, we decided to validate our results with the opinion of a general audience. For this purpose, a sample of produced riddles was generated and deployed to the Figure-Eight¹ crowdsourcing platform, in a job called *Adivinhas em Português* (Riddles in Portuguese), to be answered by human judges from Portugal and Brazil, paid for this purpose. This section describes the assessed aspects, how the sample was generated, and discusses the obtained results.

4.1 Assessed Aspects

Since our goal was to produce novel riddles, ideally with humour value, both novelty and humour potential had to be assessed. We also wanted to

¹<https://www.figure-eight.com/>

know if the audience could understand the riddle and if it actually made surprising associations. Subjects were not informed that the riddles had been produced automatically. They were just instructed to use a 5-point Likert scale for scoring the four aspects, given the descriptions in Table 5.

4.2 Validation Sample

A 300-riddle sample was created for validation, all rendered with the default presentation mode (see section 3.2.2) and randomly selected from the top-150 riddles by each model, ranked roughly based on the commonality and representativeness of their features. Our intuition was that using more common words in the riddle would improve its understanding. So, the higher the frequency of their features in the CETEMPúblico (Rocha and Santos, 2000) corpus, the higher the rank. Yet, features should be as representative as possible of the concept. For instance, asking for the hyponym of a word with too many hyponyms (e.g., person, plant, instrument) decreases solvability, an aspect considered by Labutov and Lipson (2012). So, the rank was penalised according to the number of other words related the same way as the concept is to the used features.

4.3 Results

Each riddle had the four aspects rated by three subjects, enabling us to compute the agreement on answers we knew to be subjective. Table 6 shows the global results of this validation. Considering the sample and rated aspects, it shows the proportion of answers with scores from 1 to 5, three statistical measures and the judge agreement. The latter confirms the subjectivity involved in this validation. Curiously, humour potential is where agreement is higher, due to its lower average rating. Only 6% of the riddles clearly made the subject laugh, but subjects also say that more than 30% could make someone else laugh. On the other hand, 30% of the riddles would make no one laugh. We are not completely satisfied with the results on this aspect but, given the underlying subjectivity, we still think they are interesting and provide a baseline for further iterations of SECO.

On average, interpretation was accessible, which means that most riddles could be understood, but only after a second reading. Both surprise and, especially, novelty were rated higher than interpretation. This means that the riddles use some unexpected associations and the majority is

Model	Initial concept	Features	Description	Adaptations	Riddle
NC	primeiro ministro primeiro sinistro (used)	primeiro synonymOf inicial sinistro synonymOf mau	<i>o que é inicial</i> <i>o que é mau</i>	A5, A7	<i>O que significa primeiro sinistro?</i> <i>O que é inicial e mau.</i>
[en]	prime minister prime sinister	prime, initial sinister, bad	what is early what is bad		What does prime sinister mean? What is initial and bad.
RW	eleitoral	eleito synonymOf escolhido oral saidAbout boca	<i>o que é escolhido</i> <i>uma boca</i>	A1, A2, A3	<i>O que significa eleitoral?</i> <i>Uma boca escolhida.</i>
[en]	electoral	elected, chosen oral, mouth	a mouth what is chosen		What does electoral mean? A chosen mouth.
NB	barrenar barrenal (used)	sala hypernymOf bar renal saidAbout rim	<i>uma sala</i> <i>um rim</i>	A4	<i>O que significa barrenar?</i> <i>Uma sala que é um rim.</i>
[en]	drill bar+renal	room, bar renal, kidney	a room a kidney		What does 'bar+renal' mean? A room that is a kidney.
PA	bombeiro	bom antonymOf mau	<i>bombeiro</i>	–	<i>Qual o contrário de bombeiro? maubeiro.</i>
[en]	fireman	good (=bom), bad			What is the opposite of fireman? 'badbeiro'
AB	procurador	procura antonymOf oferta dor antonymOf alegria	<i>procurador</i>	–	<i>Qual o contrário de procurador?</i> <i>oferta-alegria.</i>
[en]	procurator	demand (=procura), supply pain (=dor), joy			What is the opposite of procurator (demand-pain)? supply-joy.

Table 4: Examples for each riddle generation model

Interpretation		
1	Hard	the riddle is impossible to understand, no matter how you try;
3	Accessible	after reading it more than once and thinking about it, the riddle can be understood;
5	Easy	the riddle was understood the first time you read it.
Surprise		
1	Too obvious	the riddle is too basic or literal;
3	Somewhat unexpected	the riddle makes some interesting associations;
5	Unexpected	the riddle makes unexpected associations, you had never thought of it and would never think of.
Novelty		
1	Known	the riddle is already known by you;
3	Familiar	the riddle sounds like some other riddle you know;
5	Novel	the riddle is unknown and different from those you know.
Humour potential		
1	None	the riddle will make no one laugh;
3	Some	the riddle did not make you laugh, but could make someone laugh (child or adult);
5	Great	the riddle made you laugh and has a great humour potential.

Table 5: Assessed aspects and their description

effectively novel and previously unknown by the subjects, which was one of our goals.

Table 7 has a selection of riddles. A rough English translation was included for each but, in most cases, it hardly transmits the actual meaning in Portuguese. The selection highlights riddles that stand out in some aspect, including riddles with high novelty (1–3) and humour potential (4–6). Some riddles with low interpretation (7 and others not in the table) are longer than the average, which may explain this rate, but for others we have found no explanation besides, possibly, laziness (8, 9). To stress the underlying subjectivity, we included riddles where subjects highly disagreed on the hu-

mour potential (10–14), for which the three given ratings on this aspect are shown. One of them (10) is also one of the few with very low novelty. This is a possible cause of disagreement, as many people will not find a joke so funny once they heard it for the first time. On the other hand, this shows that SECO may generate a minority of familiar riddles, originally created by humans. The remaining riddles in the table are a personal selection with humour rated higher than average.

	Score	Interp	Surp	Novel	Humour
	1	11.8%	8.0%	6.7%	28.7%
	2	20.4%	10.7%	12.4%	24.6%
	3	23.7%	29.3%	24.8%	25.0%
	4	22.6%	33.0%	25.1%	15.7%
	5	21.6%	19.0%	31.0%	6.1%
	Mean (\bar{x})	3.2±1.2	3.4±1.1	3.6±1.2	2.5±1.2
	Mode (Mo)	3	4	5	1
	Median (Me)	3	4	4	2
	Agreement	54%	59%	56%	62%

Table 6: Global results of human validation

On a side note, 42% of the answers were from Portugal and 58% from Brazil. The average score for surprise and humour was the same for both countries and, although Portuguese judges rated interpretation and novelty 0.2 points higher, standard deviations are large.

Table 8 organises the global results according to the generation model. Due both to the high standard deviations and because we are using Likert scales, it is focused on the mode and median. It is clear that not all models result in riddles with the same quality. AB riddles seem to be the easiest to understand, the lowest surprise is for RW, and the highest novelty is for NB, PA and AB. On humour, two models should be highlighted for producing riddles with higher average potential: RC and AB. Yet, so far we have no better explanation than a higher acceptance of this kind of riddles as

#	Interp	Surp	Nov	Hum	Riddle
1	4.0	2.0	5.0	2.0	<i>Qual é o contrário de dialiso? noite-áspera.</i> (What is the opposite of dialyse (day+flat)? night-rough)
2	2.0	4.0	5.0	1.0	<i>O que significa numero? o que é descoberto e simples.</i> (What does number (naked+mere) mean? what is uncovered and simple.)
3	2.3	4.7	5.0	3.0	<i>Qual é o contrário de prevenir? prevandar.</i> (What is the opposite of to prevent ('preven'+go)? 'prev'+come.)
4	5.0	2.7	3.3	4.7	<i>Qual é o contrário de frequente? frefrio.</i> (What is the opposite of frequent ('fre'+hot)? 'fre'+cold.)
5	4.3	2.7	3.0	4.7	<i>Qual é o contrário de atropelado? alvo-vestido.</i> (What is the opposite of ran over (dark+naked)? white-dressed.)
6	4.7	2.7	3.3	4.7	<i>Qual é o contrário de malbarato? bem-caru.</i> (What is the opposite of waste (bad+cheap)? quite-expensive.)
7	1.7	4.3	4.0	1.0	<i>O que significa enfiamento? a finalidade de uma agulheta que é uma parte de rosto.</i> (What does threading (to thread/stick+chin) mean? the purpose of a needle that is a part of the face.)
8	1.3	4.0	4.7	1.3	<i>Qual é o contrário de pagamento seco? sinal verde.</i> (What is the opposite of dry payment? green light (sign).)
9	1.7	3.3	3.7	1.3	<i>Qual é o contrário de reclamar? reclodiar.</i> (What is the opposite of to complain ('rec'+to love)? 'rec'+to hate.)
10	5.0	1.0	1.7	[4, 5, 1]	<i>Qual é o contrário de pormenor? pormaio.</i> (What is the opposite of pormenor ('por'+smaller)? 'por'+bigger.)
11	3.0	3.7	4.3	[5, 2, 1]	<i>Que resulta do cruzamento entre uma unidade e um indiferente? pé-frio.</i> (What do you get when you cross a unit and an indifferent? cold feet.)
12	4.0	4.0	4.0	[4, 1, 5]	<i>Qual é o contrário de virtual? irtual.</i> (What is the opposite of virtual (come+'tual')? go+'tual'.)
13	3.7	4.7	4.3	[5, 3, 1]	<i>Qual é o contrário de bemole? mal-duro.</i> (What is the opposite of flat (note, well+soft)? bad-hard.)
14	2.7	3.3	3.3	[1, 5, 2]	<i>O que significa abalesto? um prolongamento que é ligeiro.</i> (What does flap+agile mean? an extension that is light.)
15	3.3	3.7	3.7	4.0	<i>Que resulta do cruzamento entre um sentido e o que é consistente? vista grossa.</i> (What do you get when you cross a sense with what is consistent? thick sight.)
16	4.3	3.3	4.3	3.3	<i>O que significa cãotributo? um pagamento canino.</i> (What does 'dogtribute' mean? a canine payment.)
17	4.3	3.3	3.3	4.0	<i>O que significa pronto forte? o que é imediato e não é/está fraco.</i> (What does prompt strong mean? what is immediate and is not weak.)
18	3.0	4.0	4.0	3.3	<i>O que significa ecoponto? um som que serve para coser.</i> (What does ecopoint (echo+point) mean? a sound for sewing.)
19	4.7	3.0	3.0	4.0	<i>Qual é o contrário de animal? anibem.</i> (What is the opposite of animal ('ani'+bad)? 'ani'+well.)
20	3.3	4.0	4.3	3.7	<i>Qual é o contrário de procurador? oferta-alegria.</i> (What is the opposite of procurator (demand+pain)? supply-joy.)

Table 7: Selection of generation riddles with their average validation scores

jokes by the majority of the subjects. For instance, the FC presentation, used by the RC model, relies on a linguistic construction common in jokes.

Aspect	Meas	Model					
		RC	NC	RW	NB	PA	AB
Interpret	Mo	4	4	3	3	5	5
	Md	3	3	3	3	3	4
Surprise	Mo	4	4	3	4	4	4
	Md	4	4	4	4	4	3
Novelty	Mo	4	4	3	5	5	5
	Md	4	4	3	4	4	4
Humour	Mo	3	1	1	1	1	3
	Md	3	2	2	2	2	3
Total in sample		32	45	50	52	59	62

Table 8: Human validation according to model

5 Conclusion

We have described six models for generating novel riddles, in Portuguese, from a knowledge-base, incorporated in the system SECO. Even though human validation confirmed the subjectivity of their appreciation, overall, subjects found the riddles accessible, surprising, and novel, though with a low humour potential, especially for riddles produced by four of those models. We also highlight that the generated riddles are in Portuguese, a relevant contribution, given that, to the best of our

knowledge, there was no such previous work of this kind on this language.

The ability of understanding and, in this specific case, using humour, are important steps towards making artificial agents more natural. In principle, the proposed models could be integrated in one of such agents. We have plans for a Twitter bot that will use them, but are still testing different ways of selecting riddles appropriate for a given context (e.g., recent news / top trends).

Furthermore, we aim at better analysing the results of the human validation including, for instance, correlations between different rated aspects. Using the same criteria for scoring riddles that humans actually use would provide useful insights, but it would require a great number of “consensual” riddles of the tackled kinds, which is highly unlikely. In further iterations of SECO, we will devise the inclusion of alternative models of riddles or humour (e.g., *What’s the difference between...*) and we plan to study different ways of ranking the produced riddles automatically, not only based on commonality and representativeness, but also on humour-relevant features (e.g., incongruity, presence of taboo words).

References

- Salvatore Attardo. 2008. A primer for the linguistics of humor. In Victor Raskin, editor, *The Primer of Humor Research*, chapter 3, pages 101–156. De Gruyter Mouton.
- Kim Binsted and Graeme Ritchie. 1994. An implemented model of punning riddles. In *Procs 12th National Conf. on AI*, volume 1 of AAAI '94, pages 633–638, Menlo Park, CA, USA. AAAI Press.
- Paloma Galvan, Virginia Francisco, Raquel Hervás, and Gonzalo Mendez. 2016. Riddle generation using word associations. In *Proceedings of 10th International Conference on Language Resources and Evaluation (LREC 2016)*. ELRA.
- Robert A. Georges and Alan Dundes. 1963. [Towards a structural definition of the riddle](#). *Journal of American Folklore*, 76(300):111—18.
- Hugo Gonçalves Oliveira and Ricardo Rodrigues. 2018. A set of procedures attempting at the generation of verbal humor in Portuguese. In *Proceedings of 9th International Conference on Computational Creativity*, ICCV 2018, page 307, Salamanca, Spain. ACC.
- Hugo Gonçalves Oliveira. 2018. [A survey on Portuguese lexical knowledge bases: Contents, comparison and combination](#). *Information*, 9(2):34.
- Hugo Gonçalves Oliveira, Diogo Costa, and Alexandre Pinto. 2016. One does not simply produce funny memes! – explorations on the automatic generation of Internet humor. In *Proceedings of 7th International Conference on Computational Creativity*, ICCV 2016, Paris, France.
- Hugo Gonçalves Oliveira and Ricardo Rodrigues. 2018b. Explorando a geração automática de adivinhas em português. *Linguamática*, 10(1):3–18.
- Ivan Guerrero, Ben Verhoeven, Francesco Barbieri, Pedro Martins, and Rafael Perez y Perez. 2015. TheRiddlerBot: A next step on the ladder towards creative Twitter bots. In *Proceedings of 6th International Conference on Computational Creativity*, ICCV 2015, pages 315–322, Park City, Utah. Brigham Young University.
- Igor Labutov and Hod Lipson. 2012. Humor as circuits in semantic networks. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers - Volume 2*, ACL '12, pages 150–155, Stroudsburg, PA, USA. ACL Press.
- Ruli Manurung, Graeme Ritchie, Helen Pain, Annalu Waller, Dave O'Mara, and Rolf Black. 2008. The construction of a pun generator for language skills development. *Applied AI*, 22(9):841–869.
- Rada Mihalcea and Carlo Strapparava. 2006. Learning to laugh (automatically): Computational models for humor recognition. *Computational Intelligence*, 22(2):126–142.
- Paul de Palma and E. Judith Weiner. 1992. [Riddles: Accessibility and knowledge representation](#). In *COLING 1992 Volume 4: The 15th International Conference on Computational Linguistics*.
- Pedro Pinto, João Ramalinho, and Gonçalo Castro. 2017. *O Caderno das Piadas Secas – 500 Tentativas de ter graça*. Manuscrito Editora.
- Carlos Ramisch, Silvio Cordeiro, Leonardo Zilio, Marco Idiart, and Aline Villavicencio. 2016. [How naked is the naked truth? a multilingual lexicon of nominal compound compositionality](#). In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 156–161, Berlin, Germany. ACL Press.
- Elisabete Ranchhod, Cristina Mota, and Jorge Baptista. 1999. A computational lexicon of Portuguese for automatic text parsing. In *Proceedings of SIGLEX99 Workshop: Standardizing Lexical Resources*. ACL Press.
- Ehud Reiter and Robert Dale. 2000. *Building Natural Language Generation Systems*. Cambridge University Press, New York, NY, USA.
- G. Ritchie, D. O'Mara, K. Binsted, R. Manurung, S. Coulson, B. Bergen, A. Waller, O. Stock, C. Strapparava, H. Pain, and A. Nijholt. 2006. [Computational humor](#). *IEEE Intelligent Systems*, 21:59–69.
- Paulo Alexandre Rocha and Diana Santos. 2000. CETEMPúblico: Um corpus de grandes dimensões de linguagem jornalística portuguesa. In *V Encontro para o processamento computacional da língua portuguesa escrita e falada (PROPOR 2000)*, pages 131–140, São Paulo. ICMC/USP.
- Jonas Sjöbergh and Kenji Araki. 2007. Automatically creating word-play jokes in Japanese. In *Procs. of NL-178*, pages 91–95, Nagoya, Japan.
- Oliviero Stock and Carlo Strapparava. 2006. Laughing with HAHAcronym, a computational humor system. In *Proceedings of 21st National Conference on AI - Volume 2*, AAAI'06, pages 1675–1678. AAAI Press.
- Stella E. O. Tagnin. 2005. O humor como quebra da convencionalidade. *Revista Brasileira de Linguística Aplicada*, 5(1):247–257.
- Alessandro Valitutti, Antoine Doucet, Jukka Toivanen, and Hannu Toivonen. 2016. [Computational generation and dissection of lexical replacement humor](#). *Natural Language Engineering*, 22(5):1–23.
- Diyi Yang, Alon Lavie, Chris Dyer, and Eduard Hovy. 2015. [Humor recognition and humor anchor extraction](#). In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 2367–2376, Lisbon, Portugal. ACL Press.