

Automatically Creating Word-Play Jokes in Japanese

Jonas SJÖBERGH Kenji ARAKI

Graduate School of Information Science and Technology Hokkaido University

We present a system for generating wordplay jokes in Japanese, which generates riddle style puns. By using different lexicons, different results can be achieved. Web searches are used to generate hints for the riddles. A subset of the generated riddles is evaluated manually. Using a naughty word lexicon gave funnier riddles than using normal words, though computer generated riddles in general were less funny than human generated riddles. Many computer generated riddles contain broken grammar or in other ways fail to make sense.

日本語を対象とした謎掛けの自動生成

ヨーナス・シューベルグ 荒木 健治

北海道大学大学院 情報科学研究科

我々は日本語の謎掛けを自動的に生成するシステムを提案する。本システムは謎掛けスタイルの駄洒落を生成するが、使用する辞書によって、生成される結果は様々である。謎掛けのためのヒントを生成するための情報源として Web を用いた。生成された謎掛けの一部を人手による評価を行った。システムが生成した謎掛けは人間が作成したものより評価が低かったが、下品な言葉の辞書を用いた場合の方が、普通の言葉の辞書を用いた場合よりも評価が高いという結果が得られた。生成された謎掛けの約30%は、文法的な誤りや、その他の要因のため意味を理解できないものであった。

1. Introduction

This paper presents a system that generates puns in Japanese. There have been some attempts at this earlier [9, 1, 6].

In [1] a system for generating punning riddles in Japanese is described. These riddles were generated by finding words with similar pronunciations and then constructing hints to describe the relations. These hints were generated by using semantic information provided in the lexicons used. Using a small lexicon with rich manually added semantic information gave good results while using a larger lexicon with less rich information gave less impressive results.

In contrast, we try to generate riddles without using semantic information. Instead we use the Internet to generate hints for the riddles.

There has also been some research done on generating jokes for English [3, 2, 7]. Other approaches to computational humor include trying

to recognize if a text is a joke or not [8, 5].

2. Riddle Style Puns

A very simple program for generating riddles was created. It generates *なぞなぞ*, a form of riddles mainly used to entertain children. The program only generates riddles where the answer is a wordplay joke, a pun. The puns are not very sophisticated.

When generating punning riddles, three connected things are searched for and then inserted into a fixed template. This template is “An X is an X but what kind of X is Y? Z!”. Here X and Z are two words that have similar pronunciation but different meaning, thus possibly making a pun. Y is a description that matches the meaning of Z.

An example riddle generated by the program is: “モノはモノだけど、生地がギリギリなモノはな～んだ? 安物”. A free translation to English could be (words in *italics* show the pronunciation of the

“Zは * です”
“Zは * である”
“Zは * だ”
“Zが * です”
“Zが * である”
“Zが * だ”

Table 1: *Patterns for finding hints on the Internet.*

original Japanese word): “A thing (*mono*) is a thing, but what kind of things are made of material of barely usable quality? Cheap stuff (*yasumono*).”

X and Z are found by looking through a lexicon of words, searching for words where the longer word ends with the same pronunciation as the whole short word. More sophisticated matching could of course be used, but we mainly wanted to examine whether using the Internet to generate hints for riddles was viable, so this very simple method was deemed good enough.

When Z has been selected, the hint Y is generated by searching the Internet for descriptions of the form “Zは * です” (“Z is *”) using a few different patterns (all with essentially the same meaning), see Table 1.

Web searches are also used to pick which description is most suitable. The search engine hit counts for “Z is Y” and “X is Y” are divided by the hit counts for “Z” and “X” respectively. This gives an indication of how common it is to describe Z as Y compared to describing other things as Y. The word “favorite” is for instance often found as a possible description, but is not a good hint in the riddle, since it is not specific enough. The description with the largest difference in hit count ratios is then used to make a riddle.

Since the riddle is not funny if X and Z are synonyms, a check is also done to see if the meanings are too similar. This is done by checking the word overlap of the English descriptions of X and Z in a Japanese-English dictionary, for which the EDICT [4] was used. For ambiguous words, if any translations for either X or Z overlap the pair is avoided.

When generating riddles, using different lexicons to find matching words give different results. For example, preliminary tests showed that (unsur-

Type	Score (OK)	Score (all)	Broken (%)
Random	1.4	0.3	81
Human	3.0	2.9	1
Naughty	2.3	1.7	25
EDICT	2.1	1.3	38

Table 2: *Mean scores of riddles of different types.*

prisingly) using difficult vocabulary or technical terms is rarely funny. This was also one of findings in [1], where using a lexicon with commonly occurring words gave better results than using a lexicon including obscure words.

A common conception of jokes is that naughty words are often used to make jokes. We collected a list of 400 dirty, naughty or insulting words in Japanese and used this list to generate the answer Z in one version. We also used on version with the EDICT [4] dictionary, using only words flagged as especially common. These “common” words do seem to include very many words that are rarely used, though. This is especially true for words written with a single Kanji (probably caused by these Kanjis being common in longer words).

3. Evaluation

The funniness of the riddles was evaluated by having native speakers of Japanese read and grade riddles. The scale was from 1 (not funny at all) to 5 (very funny). Since the program quite often creates riddles where the hint makes no sense or is ungrammatical, the option “I do not understand” was also available. It was also possible to skip any riddle, if for instance the user became bored and wanted to quite before reading all riddles. It was also possible to write free text comments to give feedback.

10 riddles generated using the naughty words, randomly selected from 53 such riddles were evaluated, as was 10 riddles randomly selected from 400 riddles generated using the EDICT common words. These were compared to 8 human made riddles of the same form found on the Internet. As a baseline, 10 riddles generated by taking X, Y and Z randomly from other riddles, thus making

Type	Mean (3.0)	Worst (2.6)	90% (2.4)
Random	0	0	0
Human	3	8	8
Naughty	1	2	4
EDICT	0	1	3

Table 3: *The number of riddles with scores above certain levels.*

no sense what so ever, were also included.

14 volunteers evaluated the riddles. These were native speakers of Japanese with no background in language processing, 7 men and 7 women of ages between 20 and 40. The average level of perceived funniness varied a lot between individuals, the lowest assigning a mean score of 1.0 and the highest 3.9.

In Table 2 the mean scores for the different types of riddles is shown. The percentage of judgments of the type “I do not understand” is also shown, which were normally caused by either broken Japanese in the generated riddles or incomprehensible hints. Two mean values are shown, one using only the riddles that were deemed to be understood and one including also the “I do not understand” riddles, counting these as a score of 0.

Human made riddles are deemed quite a lot funnier than the system generated riddles, but not considered very funny. Somewhat surprisingly some of the human made riddles were not understood by some of the volunteers. The riddles using naughty words were funnier than those using EDICT common words, and were also understood to a higher degree. The EDICT list still contains rare or unfamiliar words, which was mentioned in the comments from several volunteers as a point of unfunniness.

In Table 3 the number of computer generated riddles that were of similar quality as the human made riddles is shown. First, all riddles scoring higher than the average of the human made riddles. This only included one riddle from the naughty word version (and only three of the eight human made riddles). Using the score of the low-

est scoring human made riddle gave two naughty riddles and one EDICT riddle. Finally, taking any riddle achieving at least 90% of the lowest scoring human made riddle included four naughty riddles and three EDICT riddles. Some of these are available in Appendix 5..

Apart from commenting that the riddles were using unfamiliar words or rare pronunciation variants of the words, it was also often mentioned that the “X is X but what kind of X is Y? Z!” pattern does not belong in the category “jokes” in Japanese. They are considered riddles and evidently often perceived as non-overlapping with jokes, which was also mentioned in [1]. They are also considered to be entertainment for children, so some comments stated that since these are for children, grown ups (all volunteers were grown ups) do not laugh at them.

A harsh but funny comment from one volunteer was “杉は杉だけど、このプログラムの構造の杉はな〜んだ？簡単すぎ” (“A Japanese cedar tree (*sugi*) is a cedar tree, but what kind of cedar tree is this program’s cedar tree? Too simplistic (*kantan sugi*)”).

4. Discussion

Overall, the riddles are not perceived as very funny. Neither are the human made riddles, so it would perhaps be good to generate jokes of a different type instead. However, some volunteers assigned scores averaging over three, including the computer generated jokes, so at least some people find the genre entertaining.

One problem with the current program is that it uses quite simple rules for filling the riddle pattern, which fairly often leads to riddles with broken grammar. Usually because the extraction patterns are too simplistic, thus extracting too short fragments of what was written. Of course, there are also many examples of broken language use on the Internet. Especially problematic is that these are often found to be very specific as descriptions for a certain word, since it is rare to make the exact same mistake when describing other words. Thus, the program tends to believe mistakes are good descriptions.

These problems can probably easily be mitigates, by for instance giving more weight to de-

scriptions that occur many times on the Internet and by improving the extraction patterns.

A harder problem is hints that make no sense. This too is quite common in the generated riddles. This is often caused by the program finding a description that is true under special circumstances, such as: “姦淫は石打の刑である” (“Adultery is punishable by stoning”). This is not normally what Japanese people think, but the program found some biblical stories where this was told. Since no other similar sounding words seemed to be punishable by stoning, such a riddle was made. It was scored as either very unfunny or not understood by almost all volunteers. Both the word for adultery used and the word for stoning are also fairly obscure words, which a typical Japanese might not understand.

It is also fairly common to find descriptions that are the exception to the rule, such as “my favorite pervert”, which makes the program think that perverts are something you generally like. These things too can probably be mitigate by giving more weight to common descriptions, though probably not to the same extent.

A very hard problem is that often descriptions are just not funny, even though they are true. It can be too factual and make the answer too obvious, as in: “パンはパンだけど、やっぱりリバイスパンはな～んだ? ジーパン” (“Bread (*pan*) is bread, but what kind of bread is for sure Levi’s bread? Jeans (*jiipan*).”) This seems much harder to fix, especially since giving more weight to common descriptions to fix other problems tends to produce more riddles that make sense by being obvious.

Finally, a problem that seems easy to fix is that the lexicons used contain words and pronunciation variants that are obscure, which ruins jokes they appear in. Changing to a lexicon with more common words or perhaps extracting words with potential for humor from corpora of jokes would probably make this problem less severe.

5. Conclusions and future work

The lexicon used seems to have a large effect. Creating a “funny word” lexicon would likely be a good idea to improve the results. We will try to extract such lists from corpora with examples of jokes. Since these tend to be very small, thus not

giving a large enough vocabulary, it could also be useful to extend these with similar words. If for example animal nouns are found to be common in jokes, adding other animals would perhaps be a reasonable way to increase the vocabulary.

Generating hints for the riddles using the Internet instead of semantic information seems feasible, though our approach was maybe too simplistic and thus made quite many mistakes. Giving more weight to common things will likely remove many of the grammatical mistakes, but might lead to more mundane (i.e. boring) results.

Word play riddles do not seem to be considered very funny, though some of the evaluation volunteers seemed to enjoy them. Possibly other types of word play jokes that are considered less childish could be generated.

Acknowledgements

We would like to thank Mayumi Suemura and Dai Hasegawa from our lab for helping out with some translations of Japanese.

References

- [1] Kim Binsted and Osamu Takizawa. 1998. BOKE: A Japanese punning riddle generator. *Journal of the Japanese Society for Artificial Intelligence*, 13(6):920–927.
- [2] Kim Binsted, Benjamin Bergen, and Justin McKay. 2003. Pun and non-pun humour in second-language learning. In *Workshop Proceedings of CHI 2003*, Fort Lauderdale, Florida.
- [3] Kim Binsted. 1996. *Machine Humour: An Implemented Model of Puns*. Ph.D. thesis, University of Edinburgh, Edinburgh, United Kingdom.
- [4] Jim Breen. 1995. Building an electronic Japanese-English dictionary. In *Japanese Studies Association of Australia Conference*, Brisbane, Australia.
- [5] Rada Mihalcea and Carlo Strapparava. 2005. Making computers laugh: Investigations in automatic humor recognition. In *Proceedings of HLT/EMNLP*, Vancouver, Canada.
- [6] Jonas Sjöbergh. 2006. Vulgarities are fucking funny, or at least make things a little bit funnier. Technical Report TRITA-CSC-TCS 2006:4, School of Computer Science and Com-

munication, the Royal Institute of Technology, Stockholm, Sweden.

- [7] Jeff Stark, Kim Binsted, and Benjamin Bergen. 2005. Disjuncter selection for one-line jokes. In *Proceedings of INTETAIN 2005*, pages 174–182, Madonna di Campiglio, Italy.
- [8] Julia Taylor and Lawrence Mazlack. 2005. Toward computational recognition of humorous intent. In *Proceedings of Cognitive Science Conference 2005 (CogSci 2005)*, pages 2166–2171, Stresa, Italy.
- [9] Toshihiko Yokogawa. 2001. Generation of Japanese puns based on similarity of articulation. In *Proceedings of IFSA/NAFIPS 2001*, Vancouver, Canada.

A Successful Riddles

Here are some of the computer generated riddles that were scored as at least 90% as funny as the human made riddles. The mean score of the riddles and which lexicon was used to generate them is also presented. Since these are puns and thus very hard to translate in a funny way, the English translation is only given to show the meaning of the riddle. It is not translated to be funny or to preserve the word play.

- ガイはガイだけど、多いガイはな～んだ？ 圏外 (“A guy (*gai*) is a guy but what kind of guy do you often run into? No cell phone coverage (*kengai*)”, EDICT, score 2.5)
- モーションはモーションだけど、効果的なモーションはな～んだ？ プロモーション (“Motion (*mooshon*) is motion, but what kind of motion is successful motion? Promotion (*puro-mooshon*)”, EDICT, score 2.6)
- サイはサイだけど、不足がちなサイはな～んだ？ 野菜 (“A rhinoceros (*sai*) is a rhinoceros, but what kind of rhinoceros is usually insufficient? Vegetables (*yasai*)”, EDICT, score 2.6)
- 高尚は高尚だけど、許されないことの高尚はな～んだ？ 婚外交渉 (“Refinedness (*koushou*) is refinedness, but what kind of refinedness is unforgivable? Extramarital adventures (*kon-gai koushou*)”, Naughty, score 2.7)
- モノはモノだけど、増えて困りますモノはな～んだ？ ばか者 (“A thing (*mono*) is a thing, but

what kind of thing is bothersome if they increase? Idiots (*bakamono*)”, Naughty, score 3.2)