

A Complete and Modestly Funny System for Generating and Performing Japanese Stand-Up Comedy

Jonas Sjöbergh

Hokkaido University

js@media.eng.hokudai.ac.jp

Kenji Araki

Hokkaido University

paraki@media.eng.hokudai.ac.jp

Abstract

We present a complete system that generates Japanese stand-up comedy. Different modules generating different types of jokes are tied together into a performance where all jokes are connected in some way to the other jokes. The script is converted to speech and two robots perform the comedy routine. Evaluations show that the performances are perceived as funny by many, almost half the evaluation scores for the total impression were 4 or 5 (top score).

1 Introduction

When it comes to computer processing of humor two main areas exist, humor recognition and humor generation (Binsted et al., 2006). This paper falls under generation. We present a system that automatically creates short stand-up comedy like performances. Most generation systems only generate simple types of jokes, by themselves. There are few systems generating complete comic shows. Our system combines several different methods for generating quite simple jokes and then combines these into one short performance made for two performers. This is then automatically converted into speech audio, and presented by two small robots.

The performances generated are in Japanese, and similar to Japanese *manzai*, a form of stand up comedy. *Manzai* is generally performed by two comedians, one straight-man (*tsukkomi*) and one funny man (*boke*). *Boke* misunderstands or says stupid things, and *tsukkomi* has to berate or cor-

rect, for instance by exclaiming “Idiot!” and hitting *boke* on the head.

2 System

Several components are combined to produce short comedy performances. We first give an overview of the system and then explanations of the components. Though the system only generates jokes in Japanese, for ease of understanding examples in English similar to the Japanese original jokes are used in the explanations.

2.1 Overall System

First a script for the performance is generated. It starts with an introduction like “Hi, we are two not very proficient joke robots, please listen to our performance.”, simply selected from a short list. Next, one robot uses a proverb or saying in Japanese, along the lines of “Recently my life has felt like ‘jack of all trades, master of none’, you know.” The other robot then makes a vulgar joke by modifying this, perhaps like “For me it has been more like ‘jacking off all trades, masturbate none’, I must say.”. The way of berating your stupid partner common in Japanese comedy has been incorporated in the system. After the vulgar joke above, the first (*tsukkomi*) robot says a phrase from a list of fairly generic put-down phrases, like “What the hell are you saying?”.

Then the *boke* robot tells a joke from a database of wordplay jokes, selected from those with one noun already in the script. So in the example above, perhaps: “Speaking of ‘life’ [mentioned before], ‘shotgun wedding: a case of wife or death’ comes to mind.” Again followed by a put-down by the *tsukkomi* robot.

Next comes a simple punning riddle generator. A noun already used that also sounds like a rude

© 2008. Licensed under the *Creative Commons Attribution-Noncommercial-Share Alike 3.0 Unported license* (<http://creativecommons.org/licenses/by-nc-sa/3.0/>). Some rights reserved.

word is selected and a riddle is created. The riddle jokes are quite weak, similar to: “speaking’ [used earlier] is ’speaking’, but what is a naughty kind of speaking?”, “What?”, “Spanking the monkey!” (“speak” sounds like “spank” and “spanking the monkey” is a euphemism). Again followed by a put-down, “Idiot! Would you please stop.”.

Finally, one more joke from the database and another put-down are used. The robots then close with “Thank you, the end.” or similar. All the lines are then converted to speech using a text-to-speech tool. The audio files are then input into two small robots, that perform the routine.

2.2 Proverb Jokes

The proverb joke module has a list of almost 1,000 proverbs and sayings in Japanese. These were collected by simply downloading a few lists of Japanese proverbs. Since many of these are quite rare and thus people in general would not understand them or any joke based on them, rare proverbs are filtered out automatically. This is done by simply searching the web for the proverb verbatim. If it occurs more times than some threshold value (currently arbitrarily set to 50) it is considered common and can be used to make jokes, starting with a generic statement like “Recently my life has felt like <proverb>”.

To make a joke, a proverb is twisted into a new variant by changing words to similar sounding dirty words instead. The dirty words are taken from a collection a few hundred dirty words in Japanese. These have been grouped into three categories, sex related, feces related, and insults. Words can belong to several or all of these groups (e.g. “asshole”) and are then present in all groups.

A dirty variant of a proverb has to contain at least two new words, and these must be of the same type, and they must also sound reasonably similar to the words they replace. This is determined using a table of which sounds are how similar in Japanese, which is almost the same as the one used in (Takizawa et al., 1996). Since the same character in Japanese can have several different readings, we use a standard morphological analyzer for Japanese called ChaSen (Matsumoto et al., 1997) to get the pronunciation for the words. This leads to some problems, since the analyzer does not work all that well on proverbs (often using rare grammatical constructions, words, etc.), nor on dirty words (often missing from ChaSen’s

lexicons). If there are more than one way to change a proverb into a new variant, one is selected at random. The joke is then presented as described in the overview section, i.e. one robot saying the original proverb and the other saying the variant.

2.3 Riddle Jokes

There have been a few systems that generate word play riddles (Binsted, 1996; Binsted and Takizawa, 1998) and our module is not very innovative, it follows the same basic ideas. First, the script that has been created so far is run through the ChaSen morphological analyzer also used earlier. Nouns and their pronunciations are then checked against the collection of dirty words to see if there are any dirty words with similar pronunciation. A random noun sounding similar to a dirty word is then used.

The riddle is built with this noun and the corresponding dirty word using a simple pattern. The *boke* robot says :“A <noun> is a <noun>, but what kind of <noun> is <hint>?” followed by: “What?”, and the answer: “<Dirty word>”. The most difficult part is finding a hint that describes the dirty word in a good but not too obvious way without also being a good description of the original word. Hints are generated by searching the Internet for phrases like “a <dirty word> is <hint>.” Things found in this way are then assumed to be reasonable descriptions of the dirty word (often not true, unfortunately), and are then checked to see if they are also often used for the original word. This is done by checking the co-occurrences of the hint and the original noun, and the hint and the dirty word, also using web frequencies. The log-likelihood ratios are then compared, and if the hint is more closely connected to the dirty word it is used. There is also a short stop list of hints that are very common but useless, such as Japanese particles similar to the word “exist”.

Since the dirty words in our collection are not that common on the Internet, it happens that no usable hints are found at all. In such cases a simple hint meaning “naughty”, “rude”, or “dirty”, is used for sex related words, insults, and feces related words respectively. It is also happens that no noun used in the script sounds similar to a dirty word. Currently, for such cases, the whole script is abandoned and the system starts over.

2.4 Database of Puns

We automatically collected a database of word play jokes in Japanese, using a few seed jokes. If

for instance a seed joke occurred in an HTML list on the Internet, all other list items were taken as jokes too. The database consists of almost 2,200 jokes, mostly very weak word play jokes, though some are perceived as quite funny by many people. The jokes are often written using contractions (e.g. “dontcha”), dialectal pronunciation instead of standard orthography, strange punctuation or choice of alphabets etc. This causes problems for the morphological analyzer, leading to errors.

When a joke from the database is needed, all the nouns from the script up until this point are extracted as above. A joke from the database containing at least one of these is then selected and presented along the lines of “Speaking of <noun>, this reminds me of <joke with noun>”.

2.5 Put-Downs and Come-Backs

We asked an amateur comedian to write a short list of generic put-down phrases, giving things like “Ha ha, very funny”, “What the hell are you talking about?”, “Idiot”, “That is not what I meant”, and similar. Put-downs are drawn at random from the list, excluding any phrase already used.

For database jokes, two other put-downs are also possible. There is a simple web frequency check to see if the joke is old. Any joke occurring more than 20 times on the Internet is currently considered “Old!”. Jokes that are not old can instead get the “Stupid foreigner!” put-down (quite common in Japanese comedy). This is used on jokes with words written either in English letters or katakana letters. Katakana is mainly used for foreign loan words in Japanese, but is also other things (similar perhaps to using upper case in English), which leads to some errors.

For some put-downs it is also possible for the *boke* robot to make a come-back. When possible this is also added to the script. For instance, when the *tsukkomi* robot says “Old!” it goes on to say for example: “By the way, how is the new apartment you moved into?”, and the *boke* robot replies with the phrase used on him, “Old!”.

2.6 Robots

The script is converted into audio for the robots using the AquesTalk¹ text-to-speech system, and the robots are given different synthetic voices. The text-to-speech conversion works fairly well, but sometimes the speech is hard to understand.

¹<http://www.a-quest.com/aquestal/>



Figure 1: The robots used.

The two robots used in the performances are both Robovie-i robots, see Figure 1, one blue and one gold. The Robovie-i can move its legs and lean its body sideways. It has a small speaker attached, to produce the speech. This is the weakest link in the system so far, since the speaker is quite weak. The sound quality is not great, and the volume is low. This is also compounded by the text-to-speech system output sometimes being quite hard to understand to begin with, and also by the generated jokes sometimes being incomprehensible. The main merits of the Robovie-i are that it is easily programmable, cheap, and cute. The robots did not move very much. They walked a little bit forward and bowed during the introduction, then remained stationary, leaning their torsos a little to one side when speaking.

3 Evaluation

We generated two scripts and had the robots perform them for evaluators. Script 1 was shown first, then a short questionnaire for script 1 was filled out, then script 2 was performed and another questionnaire filled out. The impression of each whole performance was rated from 1 (not funny) to 5 (funny). Each individual joke was also rated.

Evaluators were found by going to a student cafeteria and offering chocolate for participating. Since the speech from the robot was a bit difficult to understand it was sometimes very difficult to hear some jokes when there was a lot of background noise. The evaluators were also given the script in written form after they had watched the performance and could thus read any parts they did not hear before evaluating the funniness. 33 evaluators took part in the evaluations. How funny the jokes are thought to be of course varies a lot from person to person. The highest and lowest means of an evaluator were 4.2 and 1.2 respectively. The results are shown in Tables 1 and 2.

Table 1 shows the overall impression of the scripts, 3.3 on a scale from 1 to 5. Joke genera-

	Script 1	Script 2	Both
Score	3.4 (0.9)	3.2 (1.0)	3.3 (1.0)
4 or 5	16 (48%)	14 (42%)	30 (45%)

Table 1: Mean (and standard deviation) evaluation scores and the number of 4s or 5s assigned, for the total impression of the two evaluated scripts.

tion systems tend to get fairly low scores, so we believe a score of over 3 is good. What meaning evaluators put into a 3 on a scale from 1 to 5 is hard to estimate, but many seemed to enjoy the performances. It was also not uncommon to laugh a lot during the performance and still rate everything as 1 or 2 so, for some, laughter does not equal funny. A score of 4 or more should reasonably indicate a funny joke. For the total impression of the performances, 30 scores (of 66) were either a 4 or a 5, so almost half of the evaluators though it was funny in this sense. We believe this is a good result, considering that individual tastes in humor vary a lot.

In Table 2 the scores of the individual jokes are shown. It seems that the proverb jokes are of about the same level as the human made jokes from the Internet. The riddle jokes lag a little behind, as does the come-back joke that was included.

While the system makes mistakes, joke generation seems rather robust to errors. Since the robots are supposed to say stupid things anyway, if they do so by mistake instead of on purpose it can still be funny. There were comments from evaluators about mistakes that they disliked too, though: “This put-down is inappropriate for that joke”, “They should bow while saying thank you, not after.”, “The dirty jokes are too direct, subtle is funnier”.

The biggest problem was the robot speakers. This should be fairly easy to fix. The other problems stem mainly from the generated jokes not being overly funny, which seems harder to deal with.

4 Conclusions

We have implemented a complete system for automatically generating and performing short stand-up comedy routines in Japanese. Different modules generate different types of jokes then tied together so that the jokes used have something in common. This is then converted to speech and uploaded into two robots that perform the comedy.

In the evaluation, the performances were rated

	Score	4 or 5
Proverb 1	2.6 (1.2)	9 (27%)
Proverb 2	3.0 (1.0)	11 (33%)
Proverb Avg.	2.8 (1.1)	20 (30%)
Riddle 1	2.4 (1.1)	4 (12%)
Riddle 2	2.3 (1.1)	5 (15%)
Riddle Avg.	2.4 (1.1)	9 (13%)
Comeback	2.6 (1.1)	6 (18%)
Database 1a	3.6 (1.1)	19 (57%)
Database 1b	2.5 (1.2)	6 (18%)
Database 2a	3.1 (1.1)	13 (39%)
Database 2a	2.9 (1.1)	13 (39%)
Database Avg.	3.0 (1.2)	51 (38%)

Table 2: Mean (and standard deviation) evaluation scores and the number of 4s or 5s assigned to the different jokes.

as 3.3 on a scale from 1 (not funny) to 5 (funny) and many evaluators enjoyed the performances. Almost half of the evaluation scores assigned to the total impression of the system were 4 or 5. This seems quite promising, though there are still many things that can be improved in the system.

References

- Binsted, Kim and Osamu Takizawa. 1998. BOKE: A Japanese punning riddle generator. *Journal of the Japanese Society for Artificial Intelligence*, 13(6):920–927.
- Binsted, Kim, Benjamin Bergen, Seana Coulson, Anton Nijholt, Oliviero Stock, Carlo Strapparava, Graeme Ritchie, Ruli Manurung, Helen Pain, Annalu Waller, and Dave O’Mara. 2006. Computational humor. *IEEE Intelligent Systems*, 21(2):59–69.
- Binsted, Kim. 1996. *Machine Humour: An Implemented Model of Puns*. Ph.D. thesis, University of Edinburgh, Edinburgh, United Kingdom.
- Matsumoto, Y., A. Kitauchi, T. Yamashita, Y. Hirano, O. Imaichi, and T. Imamura. 1997. Japanese morphological analysis system ChaSen manual. Technical Report NAIST-IS-TR97007, NAIST.
- Takizawa, Osamu, Masuzo Yanagida, Akira Ito, and Hitoshi Isahara. 1996. On computational processing of rhetorical expressions - puns, ironies and tautologies. In *International Workshop on Computational Humor*, pages 39–52, Enschede, The Netherlands.