**Abstract**

The Japanese form of word play called *dajare* was investigated in the light of the concurrent multiple word activation mechanism proposed for spoken-word recognition. Analyses of a *dajare* database revealed distinct types of punning strategies, each of which can be seen as reflecting the activation mechanism. In the present study spontaneous conversations, compiled from a live Tokyo radio talk show, provided the *dajare* evidence. These results in spontaneous Japanese again confirm the activation predictions, suggesting that *dajare* is not a low form of wit; instead, it is clever exploitation of the the natural availability of multiple words in spoken-word recognition.

**Index Terms:** *dajare*, spoken-word recognition, word play, pun, Japanese

1. Introduction

In English, it is said that “puns are the lowest form of wit”. The Japanese form of punning word play called *dajare* is also referred to as a low type of language use. This paper presents evidence, however, that the characteristics of *dajare* can be explained in the light of psycholinguistic knowledge of the process of spoken-word recognition. When a speaker produces *dajare*, he or she is making full use of how the language processing system works.

A listener’s main concern in comprehending spoken language is retrieving from the speech stream, as quickly and as accurately as possible, all and only the words that are relevant to the messages intended by speakers. In the past decades researchers in psycholinguistics have made great progress in uncovering the mechanisms of spoken-language comprehension, and it is clear that listeners actually retrieve from the speech stream many more words than the message was intended to contain.

All current mainstream models of spoken-word recognition incorporate a mechanism of concurrent multiple word activation. Speech input can momentarily activate any and all words contained in the mental lexicon which receive full or partial support from the incoming speech signal [1][2]. For example, given the auditory input *first acre*, certainly these two words will be activated one after the other; but the auditory input corresponding to these two words, the phonetic sequence [fɜstækər], in fact supports much more. At the beginning there is *far* [fɑr], and in the middle *stay* [stɛr] and *stake or steak* [steIk] and *ache* [eIk] are all added. All of these are fully supported by the signal corresponding to *first acre*. Furthermore, many partially supported words such as *farther* [fɑr’ðər] or *state* [stɛt] have their initial portions matched to parts of the input.

The important point here is that even though many of these words described above are in fact irrelevant to the messages intended by the speaker of *first acre*, they are, according to the models of spoken-word recognition, inevitably activated as word candidates as listeners hear this input. Admittedly, they are activated only momentarily, and they soon disappear [3].

However, due to the fact that words in natural languages contain very many embedded words [4], this phenomenon of irrelevant word activation is unavoidable: it is a universal mechanism across languages.

Under these circumstances, the mechanism will be operating in all language use, including that involving word play in which words are manipulated intentionally in an amusing or clever manner (rather than or in addition to the exchange of messages). Among various types of word play, puns are said to be the most popular one [5]. A pun is defined as “the use of a word in such a way as to suggest two or more meanings or different associations, or the use of two or more words of the same or nearly the same sound with different meanings, so as to produce a humorous effect” [6]. There are three categories of punning strategies based on the definition above: (1) **syllipsis** (the use of word differently in relation to two or more other words), (2) **paronomasia** (the use of words alike in sound but different in meaning), and (3) **antanaclasis** (the repetition of a word in two different senses) [7]. Although punning can be exploited both in spoken and written languages, it seems that most researchers have focused on puns in written texts such as the pun in classic Greek [8], in Japanese poetry making (kakekotoba) [9] or in Shakespeare’s literary works [10], etc.

The punning process raises an interesting question with respect to the concurrent multiple word activation mechanism described above. Since punning specifically deals with phonologically similar words such as homophonic and near-homophonic words, it concerns exactly the population that is considered to be simultaneously activated words in the models of activation. Indeed, experimental studies have demonstrated that such forms are activated [e.g., 11]. So we could view the punning process as conscious exploitation of the embedded words and alternative forms that are always available in speech recognition. Indeed, recent analyses of a *dajare* database have provided supporting evidence for this claim [12][13].

Let us look at some examples. Below, (1) is a homophonic case, while (2) is a near-homophonic case. In (1) the first *daiku* functions as a reference word and the second one as a target word. In (2), *Daisuke* is the reference word and *daisuki* the target word. These two cases are based upon phonological similarity, either exact or slightly altered.

(1) [Daiku] ґ ga utau [daiku] ґ
Carpenter SUBJ sing No. 9
Gloss: Beethoven’s No 9 sung by a carpenter.

(2) [Daisuke] ґ
Personal name like
Gloss: (We) like Daisuke.

Now consider (3) and (4). Here the relationship between the reference and the target words does not appear to be one of phonological similarity at all. However, in the light of the basic concept of concurrent word activation, these two cases can also be regarded as punning, given that the target word *komu* in (3) is embedded in the reference word *Akomu* and that the target word *sakenda* contains *sake*.
from natural conversational settings was undertaken. For the present study a new collection of spontaneous useful data would obviously come from spontaneous real speech. From the use of spoken-word recognition mechanisms, the most collected from spoken or written language. If punning arises traceable, and with no information as to whether they were collection of dajare from various sources, none of them still traceable, and with no information as to whether they were collected from spoken or written language. If punning arises from the use of spoken-word recognition mechanisms, the most useful data would obviously come from spontaneous real speech. For the present study a new collection of spontaneous dajare from natural conversational settings was undertaken.

2. Analysis

Spontaneous dajare frequently occurs when the following two conditions are met: a lively interactive conversational setting provides a listener with spontaneous auditory input, which furnishes the material for the listener to make a pun, and the participants in the situation are anxious to give an effect of humor. Thus, to test our hypothesis, we sought an appropriate source that met these conditions, so would be likely to yield dajare cases. We found the source in a popular radio program.

2.1. Dajare source

The radio program Oosawa Yuuri no yuu yuu waid, (‘Oswara Yuri’s relaxed broadcast’), a long-running (over 20 years!) daily live station in Tokyo was chosen for this study. In this program there is an interactive live talk segment called Dokumamushi Sandayuu no myuujikku purezento (‘Dokumamushi’s musical request’). In this segment, a 74-year old host (he is originally from Tokyo) visits various shops, stores and small business companies in the Kanto area. The radio host attempts to improvise lively interactions with people in each visit, and to amuse them with his own witty remarks. The important point here is that the interaction is always purely spontaneous and unprepared. Thus, dajare from these interactions must be reliably spontaneous, and accordingly suitable to test our hypotheses.

2.2. Recording and Transcribing

The section of this program was recorded daily to mp3 by Talk Master II (a portable radio with a timer) between August 2009 and April 2010. The recorded audio mp3 files were transferred to Adobe Audition 1.5 on a PC. The recorded materials were monitored by the author and whenever a witty remark was found, this portion was saved as a separate audio file and was transcribed in Japanese.

2.3. Data Analysis

Once all the witty remarks were transcribed, they were classified using the criteria employed in the earlier studies [12], [13] into three categories: (a) Homophones, (b) Mutations, (c) Embeddings. Mutations were further classified as segmental or durational. Examples of the interactions between the host (H) and the guests (G) are shown below. In the examples, […] R indicates a reference word, the intended utterance which functions as an auditory signal, while […] T indicates a target word, i.e. the pun which is activated by the reference word.

2.3.1. Homophones

In this interaction, the first hoogaku (‘law’) uttered by G functions as a reference word, which activated the second hoogaku (‘Japanese style music’) as a target word in H.

H: Daigaku de nani yatten dai?
University LOCATIVE what did COPULA
G: [Hoogaku] desu.
Law COPULA
Gloss: (I major in) law.
Law department Oh this also Japanese style music CO
Gloss: Law department? This is also Japanese style music.

2.3.2. Mutations

In Japanese a mutation can be realized by either a segmental or a durational change. The segmental changes can be vocalic, consonantal, or moraic. The durational changes can be lengthening or shortening, of either vowels or consonants. For reasons of space only consonantal examples are listed here.

[Consonantal change]
G: [Kigan] desu yo.
Near-sighted COPULA PART
Gloss: (I am) near-sighted.
H: Kigan? [kikan], no yoo na kao shite.
Near-sighted kumquat PART look face do
Gloss: Near-sighted? Your face looks like a kumquat.

[Shortening]
Here the conversation concerns party-going by a group of students, only one of whom is female:
H: Kanojo [toriko] da ro?
Her take COPULA PART
Gloss: Everyone tries to bring her[to the party], right?
G: Minna [toriko], desu.
Everyone COPULA PART
Gloss: Everyone is captivated by her.

2.3.3. Embeddings

The term Embeddings includes both cases where a target word contains a reference word, and cases where a reference word contains a target word. In this study the former is called Type I, and the latter Type II:

[Type I]
G: Neko-bu ichi-ban isogashii
cat section SUBJ most busy
Gloss: The cat section is the busiest
H: [Nekomu] no wa dame da yo.
come down PART TOPIC no copula PART
Gloss: Don’t get down (from exhaustion)!
2.4. Results

All told, 93 cases were collected from the recorded materials. The distribution of the three dajare categories, Homophones, Mutations and Embeddings was 26 (28%), 39 (42%) and 28 (30%), respectively (see Figure 1). Clearly, the three strategies were exploited to an equivalent degree; statistical analysis showed no significant difference ($\chi^2(2) = 2.18$, n.s.). It is interesting to recall that, as noted above, in the earlier analysis of the dajare database [12][13] the percentage for Embeddings was much higher. This may imply that these cases may have been overemphasized in the database collection.

Note that no tokens were observed where the Mutation and Embedding strategies were combined (imagine an English case in which the reference word import, in which port is embedded, induced a pun involving pork).

Figure 2 shows the distribution of the Type I and Type II strategies in Embeddings; there were 14 cases of each, and thus obviously no significant statistical difference ($\chi^2(1) = 0.14$, n.s.). In the previous analysis [13], Type I was more common than Type II. However, in the present analysis this was not so.

The next analyses concerned the accent pattern of target and reference words. In Tokyo Japanese, selection of lexical items is highly sensitive to pitch accent [14]. Figure 3 shows the proportion of accent matching across dajare strategies. As can be seen, accent pattern is matched well for both Homophones (81%) and Mutations (68%), but is not matched for Embeddings (29%). Statistical analysis showed a significant difference across the categories ($\chi^2(2) = 9.42, p<0.05$). Figure 4 shows the accent pattern for Embeddings separately per type. It might be expected that Type II embeddings are more likely to match in accent than Type I, in Type II the reference word is longer and the target word shorter, while in Type I the reference word is shorter and the target word longer. A shorter word mismatching in accent may still be briefly activated before the longer word with its accent pattern is fully recognized. A longer reference word would be harder to activate from a shorter target word, so there is a greater payoff for accent match in the latter, Type II, case. As Figure 4 shows, exactly this tendency was indeed observed, although the difference did not reach statistical significance (presumably due to the small sample size).

3. Discussion

The goal of the present study was to evaluate natural dajare from spontaneous Japanese conversational settings in the light of known mechanisms of spoken-word recognition. The results clearly confirm that these examples of natural word play make use of the naturally available mechanisms.

First, all of the punning strategies we have considered exploit the multiple activation of concurrent words, i.e. the basic mechanism by which words are recognized. Moreover, recall that mixed cases (e.g. Mutation and Embedding) did not occur. This is as would be predicted, since the activation of such neighbors (pork given import) would be very weak indeed. Also, the distribution of accent matching is skewed in the way one would expect from an activation explanation.

Second, the presence of embedded dajare in our corpus is notable. Traditionally, as we saw in the definition of a pun in the Introduction, the punning process is held to involve mainly homophones and mutations, i.e., phonologically similar whole words. These phonological form-based punning processes can of
course exist in any language, including in Japanese, since they rest on the fact that the vocabularies of all languages are made up of a limited set of phonemes, so that there are always many similar-sounding words. These punning processes are universal. However, our findings suggest that a third type of punning process is at least as common as the traditionally named two: embedding, at least in Japanese, is a common dajare strategy. As we shall argue below, embedding puns are particularly likely in Japanese; cross-linguistic comparisons will be needed to tell whether they are more likely in Japanese than in other languages.

In Type I embeddings, the reference word is embedded in the target word, while in Type II, the target word is embedded in the reference word. These two types can be seen as forming a denotative relationship for Type I and a connotative relationship for Type II (see the full discussion in [13]). The availability of embedded words to listeners is entirely compatible with the mechanism of concurrent multiple word activation; see [2] in which Shortlist simulations show both “catalog” and “cat” being activated by the auditory input “catalog”. In TRACE simulations, likewise, the auditory input “carpet” activates both “carpet” and “car” [1]. On the other hand, the auditory input “car” also produces temporary activation of both “car” and “carpet” [15]. Thus, both Type I and Type II Embeddings in dajare are perfectly compatible with current models such as Shortlist and TRACE.

In the analysis of the database in the previous studies, Type I outnumbered Type II, but here this tendency was not observed. We suggest that the distribution observed in the present study is the more reliable estimate of the true distribution in speech. Similar differences have been observed when all speech errors in a large radio corpus were compared with error collections based on people writing down what they had noticed [16]; both collections contained the same error types, but the distributions of the types differed. If Type I embeddings are more noticeable to listeners, it is reasonable to expect that they would be over-represented in a collection of dajare people had noticed. There is an obvious reason why Type I might be more noticeable; the reference word is shorter and has occurred first, so when the longer target word is heard, it fully contains the reference word which has been heard before. In Type II, the reference word is longer and is not fully contained in the later-occurring target.

There are three factors that render embedded dajare particularly likely in Japanese: (a) Japanese phonology and its implications for word structure, (b) Japanese mora-based segmentation and (c) the loose interpretation of double meanings in Japanese poetry making. Japanese words consist of moras, each of which could be a distinctive semantic unit. For example, sakana ‘fish’ contains the three monomoraic words: sa ‘difference’, ka ‘mosquito’, na ‘greens’; but the first two and the last two moras are also words: saka ‘slope’ and kana ‘kana’. Thus, Japanese words contain an unusually large number of embedded words due to the phonology of the language. Second, segmentation in Japanese is mora-based [17]. As we saw above, once a mora boundary is identified, new words can immediately emerge. When moras are combined together, a new word emerges. Thus, if a mora-based segmentation procedure is applied, a new word can be easily identified within a word. Finally, Kobayashi [9], reports that Japanese old poetry-making applied, a new word can be easily identified within a word. This is illustrated by the following example, where the words furusato and furu are expected to exist both furusato and furu. Thus, this tradition also activates the use of embedded dajare in Japanese (see [13]).

Finally, are there any constraints on the use of embedded dajare? If this punning strategy is so easily exploited, does this imply that nothing can inhibit the activation of embedded dajare? If this is so, Japanese spoken word recognition may be chaotic. One possible factor is that in normal conversation, prosodic structure may play a role in where embedded dajare may occur. A recent study of prosody and word boundary detection found that the accentual phrase (APR) may function as a word boundary in the recognition of spoken Japanese [18]. If so, even though embedded dajare could occur, prosody may prevent listeners from detecting the puns. Notice that under normal conversational conditions, embedded dajare is rarely detected by listeners. This may suggest that APR may be working as a kind of constraint, so that embedded dajare may not occur freely.

4. Conclusions

This study investigated the word play dajare in natural conversational settings in Japanese. Despite the fact that dajare is frequently regarded as a poor or bad pun, our results show that is in fact a very versatile type of word play in which users of the language consciously capitalize on the resources made naturally available to them by basic spoken-word recognition mechanisms.

5. Acknowledgements

The final idea of this paper was completed during my stay at the Max Planck Institute for Psycholinguistics in May, 2010. I would like to thank Anne Cutler and James McQueen for their valuable and provocative comments.

6. References