



Recognition of (Almost) Spoken Words: Evidence from Word Play in Japanese

Takashi Otake and Anne Cutler

Dokkyo University, Soka, Japan and MPI for Psycholinguistics, Nijmegen, The Netherlands
otake@dokkyo.ac.jp and anne.cutler@mpi.nl

Abstract

Current models of spoken-word recognition assume automatic activation of multiple candidate words fully or partially compatible with the speech input. We propose that listeners make use of this concurrent activation in word play such as punning. Distortion in punning should ideally involve no more than a minimal contrastive deviation between two words, namely a phoneme. Moreover, we propose that this metric of similarity does not presuppose phonemic awareness on the part of the punster. We support these claims with an analysis of modern and traditional puns in Japanese (in which phonemic awareness in language users is not encouraged by alphabetic orthography). For both data sets, the results support the predictions. Punning draws on basic processes of spoken-word recognition, common across languages.

1. Introduction

Distinguishing one word from another, and as rapidly as possible, is the listener's main concern in understanding spoken language. Current models of spoken word recognition agree that speech input can momentarily activate any and all words with full or partial support in the signal. Thus the sequence *first acre* can activate *fur* and *stay* and *steak* and *take* and *ache* as well as the intended two words. Also there can be momentary activation of words which have temporary partial support - such as *further* and *fervent*, which begin like *first*, or *state*, *stage*, *table*, *tail*. Evidence for multiple activation of candidate words, and competition between them, has been provided in numerous psycholinguistic laboratories over the past decades (see [1] for a review).

The competition process is modulated by processes that check the parse of a string as a whole; for instance, the Possible-Word Constraint [2] allows words to be rejected if the effect of accepting them would be to leave a residue which could not itself be a word. This is a simple way to get rid of spuriously embedded words such as *fur in first bus*; [st] cannot combine with [b], and alone could never form a word.

In the case of two words beginning in the same way, competition can be resolved as soon as the input provides information mismatching one competitor (but matching the other). A recent study by Soto, Sebastian and Cutler [3] showed that the effects of mismatch were equivalent for any phonemic information. In their study, listeners made lexical decisions to visually presented Spanish words preceded by spoken word fragments (cross-modal priming). Listeners accepted SARDINA as a word more rapidly if they had just heard *sardi-* rather than *sarda-* (the beginning of the real word *sardana*), PAPILLA more rapidly after *papi-* than *pati-* (*patilla*), BOFETON more rapidly after *bofe-* than *bole-* (*boletin*). In each case there was inhibition (slower responses compared with a control condition) when the spoken prime

was consistent with the competitor word. Importantly, the pattern of results was identical for mismatch in one or many features. It was the phonemic distinction which enabled listeners to reject one word in favour of another. A phoneme is, by definition, the minimal phonetic segment which can distinguish meaning, i.e. differentiate one word from another, and it is this difference which matters for listeners.

Equivalent mismatch effects at the phoneme level in no way imply that listeners explicitly represent speech input in terms of phonemes. The results are neutral as to intermediate representations in speech processing, but indicate that phonemes serve in spoken-word recognition as the minimal quanta of information which move the recognition process a step closer to resolution of inter-word competition.

Furthermore, although psycholinguistic models of spoken-word recognition are based on such findings from laboratory experiments, it should also not be concluded that the phenomena of multiple concurrent activation of word forms and inter-word competition are only observable in such controlled conditions. In contrast: language users exploit these basic processes freely in everyday life. In the present paper we exemplify this with evidence from word play.

Across cultures it is common to make puns which involve small distortions of spoken words into other words. When a sun-worshipper walks around in a T-shirt emblazoned *Sun of a Beach*, readers are invited to access a phonologically similar string. Punning is a source of linguistic information. Lagerquist [4] analysed a well-known corpus of English puns, and found that the two most common mechanisms (300 and 308 instances respectively, of 771 cases in all) involved homophony (e.g. *sun* for *son*) or mutation of a single phoneme (e.g. /i/ for /ɪ/). The remaining cases involved addition or deletion of a single speech sound (*damp* for *damm*, *fanny* for *fancy*) or transposition of sounds or words. Lagerquist concluded - in what she referred to as a "Cerface generalisation" - that successful puns involve minimal difference between the words uttered and the intended base.

This is indeed as would be predicted from spoken-word recognition theory. A successful pun should activate in the mind of the listener the intended base as well as the words actually spoken by the punster, and the greater the overlap of pun and base, the more both are likely to be activated.

However, although puns indubitably draw on speakers' and listeners' normal language processing, they are made - or perpetrated - by speakers aware of what they are doing. English speakers may choose to construct puns by altering single speech sounds simply because phonemes are accessible to English speakers' awareness. Indeed, word games such as Pig Latin in English are largely phoneme-based [5], and other phoneme-based games such as "I Spy" - in which words beginning with a given sound are sought - are popular in childhood. This pattern holds in general for languages with alphabetic orthography.



Neither alphabetic orthography nor phoneme-based word play are true of all language groups. In Japanese, for example, word games similar to "I Spy" are based not on the phoneme but on the mora [6]. Japanese is not written with an alphabet; the phonologically based kana script explicitly codes morae. Unsurprisingly, Japanese language users have little awareness of phonemes but excellent awareness of morae [7]. Morae also function in speech recognition in that their boundaries are segmentation points [8, 9].

The claim advanced above, that puns exploit basic spoken-word recognition processes, nevertheless leads to the prediction that Japanese puns should involve minimal differences between words in exactly the same way as puns in English, in any other language with alphabetic orthography, or indeed in any language at all. That is, they should tend to involve only phoneme differences. If speakers' awareness controls the structure of puns, however, Japanese puns should be based on mora substitution, and a mora substitution involving phonemic overlap (*ka* replaced by *ko* or *ta*) should be no more likely than one involving no phonemic overlap (*ka* as *to*). This is the issue investigated in the present study.

2. Analysis

2.1. Evidence from dajare

Dajare is a standard form of punning in Japanese. A web search located 257 Japanese sites devoted in full or in part to dajare. Of these, we selected an established database with the largest collection (1308 cases) for analysis.

The total of 1308 cases contained a number of duplications, and a few cases which were not readily interpretable. Further, some cases involved synonymy, or some other target-pun relationship, with no phonological overlap; and finally, 45 cases involved interplay between Japanese and English words. The rules for Japanese rendition of English loan forms are clear, but it seemed preferable simply to exclude these cases rather than to choose between the English or the Japanese rendering of the English form. In all, 378 cases were excluded for one or other of these four reasons, leaving 930 clear cases of phonologically based Japanese-Japanese dajare; Table 1 gives examples.

These 930 puns fell into three classes: homophones (131 cases), embeddings (659 cases) and mutations (140 cases).

Note that this classification appears to differ from that found in Lagerquist's [4] analysis of English puns. However, examples in her paper suggest that she variously classified embeddings as homophones or as mutations (for instance, she gives *coruscating on thin ice* as an example of the latter). She does not report her classification criteria, so that direct comparison with her results is impossible. In any case, as embedding formed far and away the largest category in our corpus, we wished to analyse it separately.

2.1.1. Homophones

14% of our analysed cases involved homophones, as in (1) and (2), in both of which two identically pronounced morphemes (respectively *koosoku*, *sensu*) are used. (Upper case forms in the literal renditions in all examples are grammatical markers.) 14% is a lower proportion than that reported for English by Lagerquist [4] (possibly - see above - because her homophone category was inflated by inclusion of embedded forms). Certainly homophone puns are common across languages.

Table 1: Examples of dajare

(1)	[Koosoku] ni [koosoku] sareru school rules by bind PASS Gloss: (We) are bound by school rules
(2)	Kono [sensu] wa [sensu] ga ii ne this folding fan TOP class SUBJ good PART Gloss: This folding fan has class
(3)	Sake ga [sake[nda]] salmon SUBJ shouted Gloss: Salmon shouted
(4)	[Geshi] no atsusa wa [ha[geshi]i] summer solstice GEN heat TOP severe Gloss: The heat of the summer solstice is severe
(5)	[Sai] wa [uru[sai]] Rhinoceros TOP noisy Gloss: (The) rhinoceros is noisy
(6)	[Baiku] wa gasorin wo [{bai} [ku]u] motor bike TOP gas OBJ twice consume Gloss: (This) motor bike consumes twice as much gas (as that one)
(7)	Doyoo wa dooyoo-suru Saturday TOP to disturb Gloss: On Saturday [something] is disturbed
(8)	Haiteku no sekai ni haitteku high tech GEN world to enter Gloss: Get into the world of high tech
(9)	Herushii ni taijuu ga herushi healthy with weight SUBJ reduce Gloss: Weight decreases healthily
(10)	Han daasu no gohan dasu half dozen GEN rice serve Gloss: Serve a half dozen rice
(11)	Yokohaba tasogare breadth twilight Gloss: Twilight in breadth (reference to song: Twilight in Yokohama)
(12)	Arufa, beeta, gumma alpha beta prefecture name Gloss: Alpha, Beta, Gumma Prefecture
(13)	Kekkon wa kekkoo marriage TOP no thank you Gloss: Marriage, no thank you
(14)	Hokkaido wa dekkaido Hokkaido TOP big Gloss: Hokkaido is big
(15)	Atami baka yo ne Atami stupid PART PART Gloss: Atami is a fool (reference to song: Watashi baka yo ne 'I am a fool')

2.1.2. Embeddings

71% of the cases involved target words embedded in other words or phrases, as in (3)-(6). Of within-word embeddings 145 were initial (3), 20 medial (4) and 58 final (5); together these totalled 223 cases (24% of the corpus). The remaining 436 embeddings (47%) occurred across word boundaries (e.g. 6). The fact that over 90% of within-word embeddings were at word edges (65% at word onset) again suggests that speakers choose puns which listeners can easily spot.

Our principal interest in embeddings was in the issue of segmentation. Recall that spoken-word recognition is modulated by the Possible-Word Constraint (PWC), in which



activated candidate words may be rejected if they leave a residue which could not itself be a phonologically possible word. If our hypothesis is correct, and puns rely on normal recognition processes, then puns which violate the PWC should be unsuccessful (because they would be difficult to recognise); hence, puns should tend not to violate the PWC. In Japanese, morae provide segmentation assistance for listeners in that their boundaries correspond to likely lexical boundaries, in the same way as rhythmic categories in other languages are used in segmentation [8, 10]. Listeners in McQueen et al.'s [9] word-spotting experiment could detect *uni* in *gyaouni* (in which *uni* is aligned with a mora boundary: *gya-o-u-ni*) and in *gyaNuni* (*gya-N-u-ni*) but not in *gyabuni* (*gya-bu-ni*). We predict that in *dajare*, embeddings of the latter type should be rare. In the examples (3)-(6), mora boundaries are respected: *sa-ke-N-da*, *ha-ge-shi-i*, *u-ru-sa-i*, *ba-i-ku-u*. In fact, strongly in support of our prediction, this turned out to be true of 100% of the embedded *dajare*, both within and across words (we did not deem it necessary to perform a statistical test!).

2.1.3. Mutations

As in Lagerquist's corpus [4], this category included additions of a sound (7)-(8), deletion of a sound (9)-(10) and distortion of one or more sounds (11)-(15).

There were 50 cases of addition, 16 of deletion, and 74 of distortion. Our prediction is that these changes will be minimal, involving most commonly only a single phoneme. In fact, all 50 of the additions involved doubling of an already present vowel (7) or consonant (8), and 15 of the deletions involved deletion of a doubled vowel (9) or consonant (10). All of these mutations resulted in an increase or decrease of the number of morae from target to pun, but involved as predicted only single-phoneme morae and moreover did not alter the nature of the segments uttered, only their number.

Since some of the 74 distortions involved more than one mora of the target word (e.g. 15), there were 82 individual mora substitutions. We analysed these as substitutions of a whole mora (as in 13, in which *o* replaces *N*, or 14 in which *de* replaces *ho*) or of only a consonant (as in 11: *ba* for *ma*) or vowel (as in 12: *gu* for *ga*). Of the 82 substitutions, both vowel and consonant were altered in 10 cases, only the vowel was altered in 19 cases, and only the consonant was altered in 53 cases. That is, there are seven times as many cases preserving either the vowel or consonant of the original mora as cases in which both vowel and consonant are replaced. Of the 53 C changes, 20 involved more than one feature and 33 a single feature (15 voicing, 10 manner, 8 place).

We attempted, for the 82 distortions, to determine whether the pattern is what would be expected by chance, by calculating the range of substitutions possible for every altered mora. For each case, we assessed the number of possible mora changes which would have preserved a consonant (V changes), preserved a vowel (C changes), or preserved neither (M changes). Japanese has five vowels and 25 possible prevocalic onsets; also all five vowels can stand alone as morae. Not all onset-vowel combinations occur in the language, however, so that there are only 103 possible morae. Three of these are special morae (the nasal coda, the geminate coda, the doubled vowel).

For the alteration of *ga-* to *gu-* in (12), the calculation was as follows. There are 103 morae, thus *ga* plus 102 others. All other vowels than /a/ could also occur after /g/, so that 4

vowel changes would be possible. Also, any of the other onsets could occur with /a/, or the /g/ could be removed to let /a/ stand alone, giving 25 possible consonant changes. The three special morae could not possibly occur word-initially, so the number of legitimate mora substitutions which did not preserve either the C or the V was $102 - (4 + 25 + 3) = 70$. By comparison, substitutions involving the mora *yu* (/ju/) allowed only two possibilities in which the consonant was preserved, since only three morae have /j/ onset. The appropriate numbers were calculated for each substitution individually.

Across the 82 altered morae, the mean number of possible substitutions involving only a C change was 21.8, the mean number involving only a V change was 3.5, and the mean number involving M change was 74.6. (Thus on average 2.1 morae were impermissible substitutions.) Converted to percentages and expressed as those percentages of 82, these counts suggest that random selection among phonologically legitimate options would yield 20.75 cases involving single-phoneme change (17.92 C, 2.83 V) and 61.25 cases involving M change. Comparison of the actual figures for C and V changes versus M change (72 to 10) with the above expected figures reveals a highly significant difference ($\chi^2 [1] = 62.66$, $p < 0.001$). Thus, puns are significantly more likely to preserve part of the substituted mora than would be expected by chance. A further comparison of the actual figures for substitution of consonant versus vowel (53 to 19) with expected figures (62.2 versus 9.8), derived by applying the ratio of 17.92:2.83 to the partial preservation total of 72, did not quite reach significance ($\chi^2 [1] = 2.9$, $p < 0.1$), suggesting that the C:V substitution imbalance is approximately what would be expected by chance.

The evidence from *dajare* strongly suggests that puns tend to resemble their targets as closely as possible; where distortions occur, they are most often distortions of only a single phoneme. Because the number of distortions in *dajare* was lower than we had expected, however, we extended our analysis to another set of puns, called *goroawase*, a traditional literary method particularly popular in the Edo period (19th century). Note that any putative phonetic sensitivity among present-day Japanese language users caused by widespread use of English words in advertising signs and brand names could hardly be postulated of speakers at that time.

2.2. Evidence from goroawase

Goroawase is another form of punning in Japanese. In its traditional form, the game is to make a comment (presumably relevant to an ongoing conversation) which is at the same time a subtle distortion of a name, a quotation from poetry or literature, or a proverb. Utagawa Kuniyoshi, an Edo period artist, produced two series of woodcuts consisting exclusively of *goroawase* jokes. Both series are based on the names of the stations (coaching inns) along the main Tokyo to Kyoto highway. The first series is made up of pictures of cats performing some action and contains all 55 station names, the second series, depicting facial expressions, contains 30 of the names. Only two of the woodcuts involve homophones and in two a mora is deleted; the remaining 81 cases involve 139 separate mora substitutions. We analysed these in the same manner as the *dajare* (see also [11] for further analyses).

Across the 139 altered morae, both vowel and consonant were altered in 44 cases, only the vowel in 22 cases, and only the consonant in 73 cases. Thus again the single-phoneme substitutions outweighed more complex substitutions. The



calculation of substitution possibilities showed that the mean number of possible substitutions involving only C change was 19.6, V change was 3.4, and M change was 69.3. Again, comparison of the actual and expected figures for C and V versus M is highly significant ($\chi^2 [1] = 140.2, p < 0.001$). Thus these puns also are significantly more likely to preserve partial morae than would be expected by chance.

The proportion of M substitutions is higher here than it was in the *dajare* set. These *goroawase* puns were of course not constructed on the fly, but were constrained by the availability of a depictable option and were further rendered more interpretable by the support of the picture in the woodcut. Nevertheless, it is striking that the evidence from *goroawase*, like the evidence from *dajare*, suggests that word substitutions, more often than would be expected by chance, preserve all of the base but a single phoneme.

3. Conclusion

Our analyses of both modern and traditional types of punning in Japanese showed that puns involving distortion resemble their targets most often in all but a single phoneme. There was no evidence in any analysis that similarity computed over morae was greater than similarity computed over phonemes. A recently published book [12], describing techniques for Japanese *dajare* for a novice readership, confirms our analysis; although the author does not use linguistic or psycholinguistic terminology, he recommends that puns be either homophonic or close in pronunciation to the underlying form. Japanese punning appears to be based on minimal difference between pun and base word, just as has been reported for English.

This is important for a number of reasons. First, it confirms our hypothesis that word play (such as punning) exploits with great efficiency the basic processes of human recognition of spoken words. Speech input activates multiple candidates fully or partially compatible with it, and a process of competition ensues in which any minimal additional information can favour one candidate and disfavour its rivals. By ensuring that puns are minimally distorted versions of their underlying bases, speakers ensure that listeners can retrieve them easily, since they are likely to be the nearest rival to the actually spoken word. Japanese is, as one might expect, no different in this respect from any other language.

Recent claims for a special status for the mora as an early "coarse coding" unit for speech perception in Japanese (see [11] for fuller critique of these claims) are thus - our second conclusion - not supported by the evidence from word play.

Third, although speakers of Japanese have greater awareness of moraic structure in spoken language than of phonemic structure, and although word play is undoubtedly something which speakers undertake with full awareness of an intention to pun, the awareness does not influence the way in which target-pun similarity is computed. Most distorted puns differ from their target by a single phoneme.

Fourth, the results from this naturalistic source of data strongly confirm empirical findings from psycholinguistic laboratories. Two types of empirical evidence are relevant. On the one hand, there is evidence collected with the word-spotting task. In this task listeners spot words embedded in nonsense contexts. The same skills are involved in listeners' recognition of puns embedded in longer words and phrases; and such embeddings in Japanese without exception respect mora boundaries and avoid the mora-splitting embeddings which proved very difficult in word-spotting studies.

On the other hand, the results confirm evidence collected with a psycholinguistic laboratory task called word reconstruction [13], which taps the very skills which listeners use to recognise distorted words in puns. In word reconstruction, listeners hear nonsense words and must recover the nearest word. In studies with this task in Japanese [11], the present authors showed that words were easier to recover when only half of a CV mora - either the vowel or the consonant - had been replaced than when the whole CV had been replaced. Thus *panorama* was easier to recover from *panozama* or *panorema* than from *panozema*. This finding clearly parallels the findings from our analyses of puns. Word play, in short, thus offers further support for current psycholinguistic models of spoken-word access processes.

4. Acknowledgements

This research was supported by grants to the first author from the International Communication Foundation, Japan and from the Japan Society for the Promotion of Science (Grant-in-Aid for Science Research (C), #11610566). We are grateful to Makiko Komada for useful tips. The *dajare* database is at <http://cactus.aist-nara.ac.jp/student/tatuo-y/dajare/index.html>. Selected *goroawase* examples may be viewed at <http://www.mpi.nl/world/persons/private/anne/cat.html>.

5. References

- [1] Frauenfelder, U.H. and Floccia, C., "The recognition of spoken words", in: *Language Comprehension, A Biological Perspective*, ed. A. Friederici (pp. 1-40), Springer, Heidelberg, 1998.
- [2] Norris, D.G., McQueen, J.M., Cutler, A., and Butterfield, S., "The possible-word constraint in the segmentation of continuous speech", *Cog. Psych.*, Vol. 34, 1997, 191-243.
- [3] Soto, S., Sebastián, N., and Cutler, A., "Segmental and suprasegmental mismatch in lexical access", *J. Mem. Lang.*, in press.
- [4] Lagerquist, L.M., "Linguistic evidence from paranomasia", *Papers from the Seventh Regional Meeting, Chicago Ling. Soc.*, 185-191, 1980.
- [5] Sherzer, J., "Play languages: With a note on ritual languages", in: *Exceptional Language and Linguistics*, eds. L.K. Obler and L. Menn (pp. 175-199), Academic Press, New York, 1982.
- [6] Katada, F., "On representation of moras: Evidence from a language game", *Ling. Inq.*, Vol. 21, 1990, 641-646.
- [7] Mann, V.A., "Phonological awareness: The role of reading experience", *Cogn.*, Vol. 24, 1986, 65-92.
- [8] Otake, T., Hatano, G., Cutler, A., and Mehler, J., "Mora or syllable? Speech segmentation in Japanese", *J. Mem. Lang.*, Vol. 32, 1993, 358-378.
- [9] McQueen, J.M., Otake, T., and Cutler, A., "Rhythmic cues and possible-word constraints in Japanese speech segmentation", *J. Mem. and Lang.*, Vol. 44, in press.
- [10] Content, A., Kearns, R.K., and Frauenfelder, U.H., "Boundaries versus onsets in syllabic segmentation", *J. Mem. and Lang.*, in press.
- [11] Cutler, A. and Otake, T., "Rhythmic categories in spoken-word recognition", *J. Mem. Lang.*, in press.
- [12] Odajima, Y., *Dajare no Ryuugi (Techniques of dajare)*. Tokyo, Koodan sya, 2000.
- [13] Ooijen, B. van, "Vowel mutability and lexical selection in English: Evidence from a word reconstruction task", *Mem. & Cogn.*, Vol. 24, 1996, 573-583.