

Lexical Selection in Action: Evidence from Spontaneous Punning

Language and Speech

56(4) 555–573

© The Author(s) 2013

Reprints and permissions:

sagepub.co.uk/journalsPermissions.nav

DOI: 10.1177/0023830913478933

las.sagepub.com



Takashi Otake

E-listening Laboratory, Japan; Max Planck Institute for Psycholinguistics, The Netherlands

Anne Cutler

Max Planck Institute for Psycholinguistics, The Netherlands;

MARCS Institute, University of Western Sydney, Australia;

Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, The Netherlands

Abstract

Analysis of a corpus of spontaneously produced Japanese puns from a single speaker over a two-year period provides a view of how a punster selects a source word for a pun and transforms it into another word for humorous effect. The pun-making process is driven by a principle of similarity: the source word should as far as possible be preserved (in terms of segmental sequence) in the pun. This renders homophones (English example: *band–banned*) the pun type of choice, with part–whole relationships of embedding (*cap–capture*), and mutations of the source word (*peas–bees*) rather less favored. Similarity also governs mutations in that single-phoneme substitutions outnumber larger changes, and in phoneme substitutions, subphonemic features tend to be preserved. The process of spontaneous punning thus applies, on line, the same similarity criteria as govern explicit similarity judgments and offline decisions about pun success (e.g., for inclusion in published collections). Finally, the process of spoken-word recognition is word-play-friendly in that it involves multiple word-form activation and competition, which, coupled with known techniques in use in difficult listening conditions, enables listeners to generate most pun types as offshoots of normal listening procedures.

Keywords

puns, Japanese, homophones, embedding, mutations

Corresponding author:

Anne Cutler, Max Planck Institute for Psycholinguistics, PO Box 310, 6500AH Nijmegen, The Netherlands.

Email: Anne.Cutler@mpi.nl

Introduction

Understanding the processes of speech perception and production is one of the major goals of psycholinguistics. The evidence gathered in pursuit of this goal comes in large part from ingeniously contrived laboratory experiments, augmented by an increasing contribution from the techniques of neuroscience. The analysis of natural speech phenomena, despite a long tradition in speech production research in particular (e.g., Al-Kisai, 1915; Meringer & Mayer, 1895), accounts for only a very small proportion of psycholinguistic undertakings.

Natural phenomena are notoriously difficult to control, and studies based on such naturalistic evidence may suffer as a result. Selection bias is an obvious danger for collections of slips of the tongue, and the reliability and validity of such collections has indeed been often questioned (see, e.g., Cutler, 1981; Ferber, 1991; Stemberger, 1985). When collections of individual errors noticed and recorded “in the wild” by one or more observers are compared with an exhaustive record of every error in a large speech corpus, different error-type distributions are revealed (Ferber, 1993), suggesting that some types of slip are more likely to attract collectors’ attention than others.

Puns and word play have also provided useful psycholinguistic insights – for instance, into the phonological knowledge underlying speech production (e.g., Katada, 1990; Sherzer, 1970). Reliability issues may hold for puns as well; collectors may tend to record puns that are memorable, more amusing or in some other way better than the average. Moreover, many puns that we encounter are brought to us by entertainment media, and will often have been carefully thought out in advance. Some collections of puns contain a mixture of spontaneous puns with puns from media sources. Other collections include or consist completely of puns from written sources, which may also be carefully crafted. Such sources can, however, yield extensive amounts of data.

Japanese, for instance, has a very rich tradition of verbal reference, allusion and word play. Many processes of this kind are conventional and intrinsic to particular genres. Thus in the classic poetic form *tanka*, *kakekotoba* (“hook word”, i.e., a verbal link) is considered basic to the form; whenever a poem referred to a pine tree (*matsu*), for example, the second meaning of *matsu* “to wait”, was effectively hooked on, so that the reader was expected to extract this additional meaning too. In the Edo period, a 19th century highpoint of Japanese culture, a conventional form of punning called *goroawase* was highly regarded; the trick here is to make a relevant comment that is at the same time a subtle distortion of a proverb or a quotation from poetry or literature. The person producing such word play was thus able to display cultural learning and linguistic facility simultaneously, and correct application of such a word-play process would in general be regarded as elegant and pleasing. It can be seen that word play in Japanese, particularly in the form of making indirect reference to another word (*share*), is not at all looked down upon but is part of cultivated language use.

Of course, in Japanese as in any other language, not all practitioners of word play achieve elegance and wit. Nonetheless, punning enjoys widespread popularity in Japan, even though the name for ordinary puns, *dajare*, is rather derogatory (it breaks down to *da* + *share*, where the *da* means “bad” or “poor”). There are many web databases of *dajare*, and conventional *dajare* practices exist, such as turning English phrases into similar-sounding Japanese. *Dajare* has attracted considerable attention from Japanese linguists (see, for instance, the many references listed by Kawahara & Shinohara, 2009).

The degree to which reliability issues influence the usefulness of a pun collection depends on the research question; collectors’ bias towards puns that work well may be an advantage if the research question concerns the factors that make puns work well! This question has indeed been the focus of much of the linguistic literature on this topic. Thus, Lagerquist (1980) analyzed

English puns from multiple sources and concluded that puns should not change the stress pattern in that language, and that consonant mutations should preserve as many distinctive features as possible. Kawahara and Shinohara (2009) made use of internet collections of Japanese puns to contrast different models of phonological similarity, concluding that source-word/pun similarity was better captured in acoustic–phonetic (i.e., perceptual) terms than by distinctive (i.e., articulatory) features.

Our own earlier analyses of Japanese word play (Cutler & Otake, 2002; Otake & Cutler, 2001) focused on the issue of the role of the mora in Japanese word recognition. The mora, a subsyllabic unit that can consist of a vowel, an onset+vowel or a syllabic coda, is the basis of the language's rhythmic structure and plays a central role in speech segmentation. However, it does not intervene in word recognition; as in any language, speech input in Japanese is processed in a fully continuous manner. Consistent with this, the literary goroawase puns from the Edo period, when mora-based poetry flowered and Western writing systems were not yet taught, showed no tendency for the mora to serve as the basic level of inter-word similarity; in contrast, puns in which one word was distorted into another by changing only a single phoneme far outweighed those with changes at the mora level (Cutler & Otake, 2002). Modern puns from internet databases showed the same pattern (Otake, 2010; Otake & Cutler, 2001).

Together these conclusions from analysis of written resources suggest that the best-working puns are those in which the pun is closest to its source word, with the smallest possible change being preferable to larger change. We can term this the Principle of Maximal Source Preservation (hereafter MaxSP).

What we cannot ascertain from the written collections, however, is whether MaxSP, the principle that makes puns work well enough to be included in internet databases, enshrined in goroawase performance, or published for the entertainment of millions, also constrains pun-making in natural conversations. That is, is the MaxSP only an evaluation criterion used to assess the quality of speakers' word play after the fact, or is it also an online constraint on speakers' lexical selection during the practice of word play? In the present report we attempt to answer this question by analyzing puns from an unusual single-speaker source of abundant spontaneous puns.

Among the flowerings of the pun form in Japanese is a live segment in a Tokyo radio show. In this segment, which has been running now for over 30 years, a performer who goes by the pseudonym Dokumamushi Sandayuu (D) visits small business offices or shops and improvises light-hearted conversation with whomever he happens to meet. The aim is purely to amuse, both his interlocutors and the radio audience. The conversations take many forms, but often begin with D asking everyday questions of those he encounters; of interest to us is that D then quite often uses the answers he receives as the basis for on-the-spot punning transformations of one word into another.

The process is in no sense subtle or unexpected. The source word for a transformation is never provided by D himself, but always by one of the ordinary folk he has surprised with his visit. The products of this effort are sometimes witty, although more often lame and/or ludicrous. To give non-Japanese-speaking readers a feeling for what we can only call the corniness of the style involved, we constructed the four English analogs in Table 1, in each of which the first utterance may be thought of as an interlocutor's answer to some question (with source word underlined), and the second utterance contains a D-style pun (underlined) on some word in that answer.

The important point for the purposes of our investigation is that the interactions are broadcast live and D has not previously met his interlocutors.¹ D's puns are therefore not prepared in advance, and no editing process selects only the most memorable or most successful examples. These puns therefore provide us with a view of the punning process in action. Moreover, they put us in the

Table 1. Constructed English-language pun examples.

-
- (1) a. I sing with a band. Why were they banned, were they that bad?
 b. I've lost my cap and need a new one. I hope you capture one!
 c. I sell plumbing supplies. That's a plum job, I guess!
 d. Dried peas are my favorite snack. Dried bees – really???
-

unusual situation of having multiple examples from the same speaker. This is the corpus we analyze in this report.

1.1 A task analysis of punning

What does the process of punning in fact involve? The punster's (self-imposed) task clearly draws on both speech perception and speech production. The first step in the process is to recognize what the interlocutor has said, monitoring this speech input for suitable source words and evaluating each such word with respect to its potential for transformation into another. If there are multiple transformation options, these can be compared for amusement potential: the goal is that the resulting utterance should elicit laughter from the audience. Thus the audience's speech perception processes must be taken into account: the selected pun must, above all, be noticeable. Once a transformation has been chosen, the task becomes speech production. The pun word has to be embedded in a plausible framework and uttered. The whole process has to run its course as rapidly as can be achieved.

In ordinary circumstances, the process of speech perception starts with acoustic–phonetic input and generates the associated meaning; the process of speech production starts with meaning and generates a phonetic form to convey it. Punning, however, brings about modifications of, and even overlaps between, these two operations. Speech production involving a pun starts with the phonetic string and contrives a contextual meaning to accommodate it, thus allowing the production proper to be launched. Speech perception looking for a punning opportunity constrains the interpretation of the input so that not only exact matches but perhaps also similar-sounding words are identified.

The constraint on speech production, that a context must be contrived for a specified word form, essentially imposes a delay on that part of the task. The constraint on speech perception is potentially less delaying. The process of speech perception in fact offers structural support for the search for similar-sounding words. Empirical studies over several decades have shown that listening to speech involves selection of the correct interpretation from among a multitude of alternatives (see Cutler, 2012, or McQueen, 2007, for reviews). An incoming speech signal activates all words that are fully compatible with any part of it, and temporarily activates words that are partially compatible. Non-identical but similar-sounding words are in principle a nuisance, to be discarded as soon as possible so that the actual message can be apprehended. However, they are there, their forms automatically activated to some extent by the input.

The putative role of the MaxSP would play out in the perceptual selection process whereby the punster chooses, from all the words in the input presented, a source word and a punning transformation of it. The automatic activation of word forms will make homophones most readily available; *mail* will also activate *male*, and vice versa. Words that are embedded in other words will also be activated; *maelstrom* and *blackmail* will also make the form equivalent to *male* available. If none of the automatically activated candidates delivers a result satisfying the punster's aspirations, then the remaining option is to change the source word into something else. Here the punster has to decide how much to modify and where to apply a modification.

Does speech perception offer the punster any room to maneuver here? Multiple simultaneous activation of alternative possible interpretations of the input are resolved, in the speech perception process, by allowing the activated candidates to compete with one another until the best-supported sequence of words ultimately triumphs. This competition has been observed in many types of experiments (e.g., McQueen, Norris, & Cutler, 1994; Vitevitch, 2002) and has been successfully captured by computational models of spoken-word recognition, both connectionist (McClelland & Elman, 1986) and Bayesian (Norris & McQueen, 2008). As additional speech input arrives, some candidates will prove to be mismatched, and the competition process allows both these instances of words that are only temporarily supported, and also words that are fully supported but less well so than their rivals, to be discarded. In normal speech perception, this all happens with ruthless efficiency, and so rapidly that we as listeners are hardly ever conscious of the competitors' fleeting presence.

Interestingly, however, recent findings have demonstrated that the criterion for discarding versus maintaining active competitors is at least to some extent under listener control. One of the ingenious psycholinguistic tools for examining speech perception in progress involves monitoring listeners' looks to a multi-option display as they hear speech. Such eyetracking studies show that speech input is indeed processed efficiently and words are rejected immediately if they are mismatched. A consequence of this is that competitor words that overlap at onset with what is being said compete for longer than competitors that overlap at offset; if *candle* is heard, both a candy and a handle will receive some looks, but the candy will get more, for longer, because mismatch between the input and the spoken form *handle* occurs early. When listening conditions are less than ideal, however, this onset/offset asymmetry is less pronounced. Just occasional crackle, as of an imperfectly tuned AM radio, can have this effect, and the crackle does not need to interrupt the target words or to coincide specifically with word onsets; even words that are themselves clearly pronounced are more likely to induce looks to offset competitors and less likely to induce looks to onset competitors if other utterances are interrupted in this way (McQueen & Huettig, 2012). Similarly, casual speech – again, elsewhere in the utterance – can have the same effect of modifying the competition patterns evident in listeners' looks to items in a display (Brouwer, Mitterer, & Huettig, 2012). Listeners allow for the possibility that their perception may be inaccurate in challenging environments, and adjust their listening tolerances accordingly. The possibility of such adjustment may also be exploitable for punning purposes.

With this discussion of the component processes of punning in mind, we analyze the type of puns we found in the D corpus, and the factors constraining the relationship between source word and pun.

2 Data collection

From the radio program *Oosawa Yuuri no yuuyuu waido* ("Osawa Yuri's relaxed broadcast"), which airs live for four and a half hours each weekday morning on the Tokyo broadcasting station TBS, we selected for analysis the interactive live talk segment *Dokumamushi Sandayuu no myuu-jikku purezento* ("Sandayuu Dokumamushi's musical request"). This segment of approximately 25 minutes airs around 10.25 a.m. each day. Using a Talk Master II portable radio with timer, we began recording the segment intermittently from August 2009, then systematically from mid-November 2009 until late August 2011. During this period all available occurrences of the segment were recorded, availability being interrupted only when the segment was off air due to vacation, the first author of this paper was away from home or (on a few occasions) a technical malfunction prevented recording. This yielded in total over 400 recordings.

Recordings were made to mp3 and the files transferred to Adobe Audition 1.5 on a PC. The recorded materials were monitored by the first author and puns (with their entire context) saved as separate audio files and transcribed in Japanese.

Puns for analysis were defined as word transformations, with a source word spoken by an interlocutor and a pun word produced by D. Thus we excluded puns where the source word, although obvious, was not part of the triggering utterance (imagine responding to a request to admire a new fireplace by saying: *It's great!*). Not every segment yielded a pun according to this definition (although some segments yielded several puns). D has a significant repertoire of strategies to produce a light-hearted listening experience, of which punning is just one; there is no way to tell whether he himself ranks it highly among his options, or even whether he feels the same way about it every time. Note, however, that it is a well-enough established feature of his interactions that many of his interlocutors also tried to produce on-the-spot puns while conversing with him. Despite the fact that D does not make a pun in every utterance or even in every interaction, our collection yielded 294 such single-word puns with a clear source word. No puns were repeated in this total yield. We also observed 62 puns from other speakers fitting the same criteria.²

By definition, our selection criteria allow three possible relationships of the pun to the source word. (1) Homophone puns, in which the source word and the pun word are identical in form; the English example (1a) in Table 1 is such a case. (2) Embedding puns, in which the source word and the pun word have a part-whole relationship. Either the source word can be inserted into a longer word to create the pun (example 1b) or the pun can be extracted from a longer source (example 1c); in either case, the form of the part is preserved in the whole. (3) Mutation puns, in which the source word and pun are not identical in form, but some part of the source has been transformed to produce the pun (as in example 1d).

Twelve transcribed and glossed illustrations from the actual corpus are listed, with their contexts, in Table 2. The initial occurrence of the source word uttered by an interlocutor (I), and the occurrence of the pun word in an utterance by D, are underlined in each example. Homophone examples 2a and 2b are, respectively, a simple reference to another feature of the interlocutor (he described himself as a law student, but was carrying a Japanese instrument: *hoogaku* means both 'law' and 'Japanese music'), and a more convoluted example (the interlocutor is taking night-school classes; D remarks that a kettle is as good as a teapot, implying that just as one can pour tea equally well from a kettle, one can gain a university degree equally well from night school, where *yakan* is both "night" and "kettle"). Examples (2c) and (2d) are insertion embeddings, (2e) and (2f) extraction embeddings. Examples (2g)–(2l) are all mutations; in (2g), the initial consonant is altered (*rikei* becomes *bikei*); in (2h) a medial consonant is altered in a four-mora compound (*kingan* becomes *kinkan*); in (2i) a vowel is changed (*hibi* to *hebi*); and in (2j) to (2l) a mora is altered to a differently structured mora, respectively an initial CV being replaced by a V (*ne-n-ki-n* to *i-n-ki-n*), a medial V being replaced by a CV (*Ta-i-wa-n* to *ta-ku-wa-n*) and a medial CV being replaced by a moraic nasal (*ka-ra-o-ke* to *ka-N-o-ke*).

3 Data analysis

We analyzed source words and their contexts, selection of pun types and pun properties (preservation of source, position of embedding or change) using non-parametric statistics (Siegel, 1956). In cases testing the significance of a distribution across two alternative categories (e.g., structure-preserving versus non-preserving puns), we used two-tailed binomial tests, and in cases involving comparison against distribution-based predictions (e.g., from the vocabulary), we used chi-squared tests. All reported z values are significant at $p < 0.001$ and all reported χ^2 values have one degree

Table 2. Twelve examples of D's puns, with context.

- (2) a. D: Daigaku de nani yatten dai? "What are you doing at university?"
I: Hoogaku desu. "law"
D: Hoogaku-bu? Aa, kore mo hoogaku da. "Law department? Or Japanese music?" [reference to an instrument carried by I]
- b. D: Kyoo wa kinyoo da. "It's Friday today."
I: Baito wa yasumi de, daigaku wa yakan nande. "I've a day off from my part time job, and my university classes are at night."
D: Ah, nibu ittenno ka. Yakan demo dobin demo iin da yo. "Oh, you go to night school. A kettle is as good as a teapot!"
- c. D: Doko ni ita no "Where were you?"
I: Suma desu. Uchi wa daijoobu datta "In Suma. Our house was all right."
D: Sumai wa daijoobu datta. "Your house was safe."
- d. I: Neko-bu ga ichi-ban isogashii "The cat section is the busiest."
D: Nekomu no wa dame da yo. "Don't collapse (from exhaustion)!"
- e. I: Karada, hanshin kikanakute nookoosoku datta no. "Since he had a stroke, half of his body doesn't work."
D: Iya, kyoo wa koosoku ga dame de ne. "Well, the highway was very crowded today."
f. I: igaito ijime-rarete imasu kedo. "Surprisingly, (I am) abused, though."
D: igaito iji ga warui desu yo. "Surprisingly, he is ill-natured, isn't he?"
g. I: Ano, rikei desu kedo "Well, I am in a science course."
D: Oh, rikei ka. "Oh, a science course."
I: Hai. "Yes."
D: Kan Naoto mo rikei da. Ore wa bikei da. "Kan Naoto [former prime minister] also took science. [But] I am esthetic."
- h. I: Kingan desu yo. "[I am] near-sighted."
D: Kingan? kinkan no yoo na kao shite. "Near-sighted? Your face is like a kumquat."
- i. D: Yononaka heiwa ni naru yoo ni inotte kureta? "Did you pray for peace in our time?"
I: Ee, hibi "Yes, every day."
D: Hibi ne. "Every day?"
I: Hibi. "Every day."
D: Ore wa hebi da kedo. "I am a snake though." [referring to a nickname of his]
- j. I: Nenkin no renraku shitekure temo ii yone. "They should tell us when we are owed a pension."
D: Soo dayo ne. 25 nen tattara. "Right, after we've been paying in for 25 years."
I: Soo iu mon deshoo. "That's the way it ought to work."
D: Ore datte haratte iru kedo, hitotsuki nanzenman haratten dakedo. (Usu!)
Nenkin ja nakute inkin ni nacchata yo. "I've been paying in a million yen every month. (Just kidding!) You could call it ringworm rather than pension."
- k. I: Otoosan wa Taiwan daigaku o sotsugyoo desu kara "As my father was a graduate of Taiwan University."
D: Taiwan daigaku "Taiwan University."
I: Sono tooji wa ne "At that time"
D: Kocchi wa takuwan daigaku. "Me, I'm from dried radish university."
- l. D: Okusan ga sukina uta ka? "Is this your wife's favorite song?"
I: Soo desu ne. "That's right."
D: E. "Really?"
I: Karaoke de yoku yarun desu. "We often sing it at karaoke."
D: Aa soo. Issho ni iku no. "Oh, I see. Do you go together?"
I: Ikimasu. "We do."
D: Izure ka Noke da na. "You will go to a coffin someday". [Reference to a proverb saying that a good husband and wife go to the same grave.]

of freedom and are significant at $p < 0.001$, except for two noted exceptions; for tests reported as insignificant, p was >0.1 . Table 3 presents overall summary statistics of the corpus.

3.1 Source-word characteristics: What word is chosen to be punned upon?

D's puns involve nouns. Across all pun types, the source words are overwhelmingly common (218) or proper (72) nouns; in our entire corpus only four of the source words are not nouns (one verb, one adjective, two adverbial ideophones).³ The puns that result show the same asymmetry, albeit somewhat less marked: 269 common (252) and proper (17) nouns, 25 other parts of speech (16 verbs, eight adjectives, one conjunction), with no tendency for non-nouns to cluster in any pun type (6.6%, 9.4%, 9.7%, respectively).

The utterances from which D chose source words were short:⁴ on average around three words (3.17 for homophones, 2.94 for each of the other two pun types), with the remaining words in the utterance also being short, comprising on average only just over four moras. (In Table 2, the source-word utterances are those spoken by I and containing the underlined word. Example (2h) in the table, *kingan desu yo*, with two additional words comprising three moras, is perhaps closest to the corpus-wide average.) Neither utterance length nor source-word length itself differed across pun types.

The Japanese lexicon has several types of etymology, of which the main ones are 'yamato' (words of native Japanese origin) versus 'kango' (words of Chinese origin); loan words (including from English) form the third largest class. The lexicon contains 34% yamato, 49% kango and 9% loan words (Kindaichi, Saeki, Oishi, & Nomura, 2002), and a token count based on 1990s magazine publications (National Institute for Japanese Language, 2006) similarly shows 36% yamato, 50% kango and 12% loan words (in each case the small residue represents mixed-origin forms; equivalent counts for spoken language are not available).

We established the etymology of all source nouns in our corpus. A comparison with the vocabulary was only possible if the 72 proper-noun source words were omitted. Overall, the common-noun source words broke down to 44% yamato, 46% kango and 10% loan words. However, as Figure 1 shows, the etymological types were not equally distributed across pun types; among homophone puns, kango words were significantly over-represented ($\chi^2 = 8.45$, $p < 0.01$). Two-thirds of homophone source words were two-syllable four-mora words (e.g., *kyoodai* "brother; dressing table"; *koojoo* "improvement, factory"; *sampai* "worship; industrial waste"), which are famously homophonic. This result suggests that D is matching his puns to the opportunities provided by the input, a question that we explore further by analyzing which types of pun occur in the corpus.

Table 3. Summary statistics of the pun corpus.

Pun type	N	% Noun		% possibility of alternative pun type			% source-pun accent match	Mean length of source (moras)
		Source	Pun	H	E	M		
H[omophone]	105	100	93.4		100	72.4	77	3.24
E[mbedding]	96	95.8	90.6	37.5		73.1	52	3.13
M[utation]	93	100	90.3	6.5	100		81	3.28
Total	294	98.6	91.5				70	

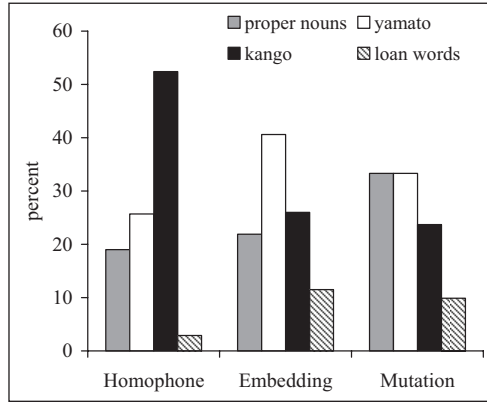


Figure 1. Etymological origin of source nouns in the pun corpus, by pun type.

3.2 Pun type: Is one type of pun preferred to another?

According to the MaxSP, the most preferred pun should retain the integrity of the source word. We would therefore predict that the most preferred pun should be a homophone. However, although homophones are not dispreferred across languages, and indeed those languages with a relatively simple phonological structure, such as Japanese, tend to contain proportionally many homophones, it is still far from the case that all words are homophonic. Around 36% of content words in Japanese written texts are potentially homophones (Nakano, 1989; Tokuhiki, 2005; no data are available for spoken language). The maximum number of homophones we would expect in our corpus would therefore be the maximum that we would expect to find in the input, namely about 36% of the content words. Of the current corpus, 35.71% (105 cases) consists of homophones. Thus, the collected corpus of content-word puns contains almost the same proportion of homophone puns as the content words in the dictionary contain homophones.

Homophones are stand-alone forms with more than one meaning, such as *match*, *band–banned* or examples (2a) and (2b) in Table 2. However, as described above, there is another way to preserve a word’s integrity in a pun, and that is to embed one word as a whole in another. Our corpus contains 96 embeddings, of which 51 are insertions of the source into a longer word (e.g., examples 2c, 2d) and 45 are extractions from a longer source word (e.g., examples 2e, 2f), an insignificant distributional asymmetry.⁵

The overall distribution of the 294 puns in the corpus is thus 105 homophones and 96 embeddings, leaving 93 cases in which there is alteration to the source word (mutations). The two preserving pun types, homophones plus embeddings, account for 201 cases, significantly outnumbering the 93 non-preserving mutations ($z = 6.36$).

3.3 Pun type: Was there actually a choice?

We analyzed all source words in terms of the opportunities they offered for alternative pun types: Was the source-word homophonic? Did it contain embedded short words (of more than a single mora) and could it have been embedded in a longer form? Would changing a single sound at any point result in another real word?

3.3.1 Mutations. The MaxSP predicts that mutations will occur when source-preserving selections are not an option. Thus the source words in the mutation set should not consist of roughly one-third of words with homophones, as in the vocabulary as a whole, but of significantly fewer such cases. Indeed, only seven of the mutations (6.5%) have source words with homophones, significantly fewer than chance would predict ($\chi^2 = 32.22$). Note that homophone patterning is not determined by the frequency of proper nouns in our corpus; in Japanese, as in other languages, names are very likely to have other meanings (*ootake* is a big bamboo, just as a *cutler* is someone who makes cutlery).

Embedding puns would, in contrast, have been possible with the mutation source words. Japanese compounding processes would effectively allow any source word to have been embedded into a longer (potentially nonce) form, and 70 mutation source words also have shorter words embedded within them (most of the remainder failed to contain an embedded form due to phonological constraints, such as gemination).

Further, the mutation source words allowed more options than those D chose: 81 of the 93 cases would have allowed other single-phoneme substitutions producing a real-word outcome (in 41 cases, more than five other outcomes).

3.3.2 Embeddings. Homophone scarcity also appears with extractions from embeddings: 42 of the 45 cases have no homophone and only three (6.6%) have a homophone, fewer than chance ($\chi^2 = 16.38$). Embeddings involving insertion, unsurprisingly given that the source words are necessarily short, pattern differently: 33 (69%) source words have homophones and 18 do not (significantly more homophones than chance would predict: $\chi^2=19.32$). In these cases the homophone counts, are, in fact, rather high: these 33 source words average 5.03 meanings each, presenting a potential embarrassment of riches for the punster.

Mutations, however, were nearly always possible with the embeddings involving insertion: 47 of the 51 source words could become another word by changing a single sound (in 40 cases, five or more other words). The one pun type that is out of line here is the embedding involving extraction; the source words in this case are necessarily long and thus much less likely to yield another word differing in only a single sound; 21 of the 45 cases have no single-substitution neighbors, although the remaining 24 do have such neighbors (in fact, more than five of them in seven of those cases).

3.3.3 Homophones. Crucially, when homophones were used as the source word for a pun, they were not chosen because no alternative was possible. Of the 105 homophones, mutating a single sound would have produced another word in 102 (one other word in eight cases, two to five other words in 14 cases, five or more other words in 80 cases). In addition, embedding the homophonic source word into a longer word would always have been possible. Embeddings involving extraction would also have often been available; of the 105 homophones, 21 source words are two moras in length (thus by definition making only a single-mora embedding possible) and eight have no embeddings due to phonological constraints, but the remaining 76 all contain an embedded shorter word.

These counts show that the options allowed by mutation outnumber source-preserving options; yet, as we saw, source-preserving options are chosen significantly more often than mutation. Thus the selection of D's source words shows a preference for homophones where possible (except perhaps when there are very many of them), and a preference for the part-whole relationship (embedding) over mutation. We now consider what constitutes source preservation versus mutation.

3.4 Preservation criteria: What may be changed and what not?

Source-word preservation can be assessed at varying levels, from less to more detailed: (a) abstract phonological structure; (b) segmental structure; (c) segmental and prosodic realization. English analogs of these alternatives are *sighted* and *fated* (same abstract phonological structure but some segments differ), *insight* and *incite* (same segmental structure but stress differs) and *sighted* and *cited* (matched at all levels). In English puns, as described in the introduction, stress change is discouraged, and the literature would lead us to predict that preservation level (c) would be more highly desired than the others.

However, is this also true of Japanese (in Tokyo, the variety represented in our corpus)? Examples illustrating the same comparisons in Japanese could be (a) *samba* and *tombo* (same phonological structure, different segments), (b) *samba* “samba” and *samba* “midwife” (same segments, but differing in accentual pattern: HLL versus LHH), and (c) *tombo* “dragonfly” and *tombo* “rake” (sharing both segments and accentual pattern, LHH). Does the (c) type (an example from our corpus) represent a more highly preferred level of preservation than the (b) type (which is in fact also an example from our corpus)?

There is reason to believe that Japanese may differ from English in this respect. The lexical prosody of Japanese is based on pitch accent. Accent placement is not free in the sense that, say, stress is in a completely free stress language. Words can be unaccented, but in words with accent there is only one accented mora, labeled high (H). If it is the word-initial mora, moras following it are labeled low (L). If it is later in the word, the first mora is labeled L, any further mora before the accent as H and any mora following the accent as L. There are thus three accent possibilities for a three-mora word (HLL, LHL, LHH) and four for a four-mora word (HLLL, LHLL, LHHL, LHHH). Note that many notional patterns in the accent system are thus impossible (e.g., *LLH, *HLH, *HLHL, *LLLH, *HHHH and so on). Moreover, not all possible patterns are equally likely. For instance, a six-mora word could in principle have any of the six accent placements HLLLLL, LHLLLL, LHHLLL, LHHHLL, LHHHHL or LHHHHH, but the first of these patterns is very unlikely (2% of six-mora words: Sugito, 1995). Initial accent in fact becomes less likely with increasing word length (66% of two-mora words are HL, 48% of three-mora words HLL, 7% of four-mora words HLLL).

Furthermore, compounding has prosodic implications in Japanese (as in English; consider compound stress in *blackbird* versus the phrase *black bird*). Length and accent pattern of the second element determine the accent pattern of a compound as a whole (Kubozono, 1998), so that accent shift may be obligatory. Our corpus contains many compound words, for which the accent pattern is constrained in this manner.

Such constraints entail that accent placement cannot be considered a matter of choice and that the vocabulary will often not allow one or another option. In particular, the strong relationship of accent to word length has implications for the source–pun pairs of our corpus. On chance alone we would expect that in puns where word length is preserved (homophones, mutations), accent might also tend to be preserved, but puns that do not preserve length (embeddings) would also be unable to preserve accent.

This is indeed the case. In 77% of homophones and 81% of mutations the accent pattern of the source word is preserved in the pun ($z = 5.46$ and 5.81 , respectively). Further, recall that the source words contain a high proportion of proper nouns (place, institution and personal names: 19% for homophones, 33% for mutations). Such words may not pattern with the vocabulary as a whole. Indeed, the accent-unmatched cases contain, respectively, 33% and 50% proper nouns, the accent-matched cases 15% and 29%. This suggests that proper nouns are less able to yield an accent-preserving pun.

In embeddings, in contrast, there is as predicted no significant matching tendency: only 52% of pairs preserve accent. In 24 of the 26 cases of an insertion where a shorter word has a different accent pattern when embedded in a longer word, the shorter word was H-initial but its originally initial syllable is not accented in the embedding (example: *genki* HLL in *genkin* LHHL). The length relationship in the vocabulary predicts such a pattern. We conclude, therefore, that accent pattern is not a criterion for preservation. The preservation of the source word is ensured in this corpus by maximizing segmental overlap, not by matching the full phonetic (segmental and prosodic) realization.

3.5 Constraints on alteration: Can anything and everything be changed?

A mutation is any pun produced by altering source-word segments (as in English dummy example 1d, or examples 2g–l). As documented, mutations are made chiefly when no homophonous meaning is available. The MaxSP predicts that such changes to the source word will be minimal, so that as much as possible of that word is preserved.

Maximal preservation entails that despite the evidence for moraic segmentation in Japanese (Cutler & Otake, 1994; Otake, Hatano, Cutler, & Mehler, 1993), and for the mora's role in lexical representations and access (Kubozono, 1989, 1990), submoraiic substitutions should outweigh moraic. Indeed, 71 of 93 cases involved a single phoneme substitution, with 22 involving more than one phoneme or a phoneme/mora addition (e.g., a long vowel replacing a short vowel, or a geminate replacing a single consonant). Since replacement even of a single phoneme creates a different mora, however, we must evaluate the statistical significance of this pattern in the light of phonological possibility. How often would random substitution of a mora lead to a result differing in only one phoneme from the source? This computation varies for each source mora and for each word position. Japanese has 103 moras, most with CV structure. However, of the five vowels of Japanese, /e/ and /i/ do not occur with some onsets, and the three "special" moras (nasal coda, consonant gemination and vowel doubling) cannot occur word-initially (and gemination also not word-finally). As in Cutler and Otake (2002), we calculated the phonologically legal substitution possibilities for each source mora individually to determine the expected ratio of minimal (single-phoneme) versus non-minimal (whole mora) changes in the 93 mutations. This calculation yielded an expected ratio of 25 (27%) minimal to 68 non-minimal changes. The observed ratio in the set is 72:21, a highly significant difference from the expectation ($\chi^2 = 120.85$). Minimal changes were predominant in all word positions (initial 68%, medial 87%, final 87%).

Further, single consonant substitutions are influenced as well by requirements of similarity at the subphonemic level. For the 53 consonant substitutions, the expected values (again calculated from phonological possibilities) would be 11 cases preserving two broad features (place, manner, voicing), 28 cases of one preserved feature and 14 cases with no features preserved. The observed values are, respectively, 36 two, 14 one and three no-feature cases, again a highly significant difference from the chance expectation ($\chi^2 = 72.46$).

Finally, we analyzed similarity in the 21 non-minimal, or mora substitutions, that is, all changes where one mora structure (e.g., V, CV) was replaced by another. This can include consonant deletion with vowel preservation (e.g., *ke* to *e* in *ooboke-ooboe*, *i* to *ni* in *imono-nimono*), an option that may be reckoned to involve higher similarity and hence to be preferred. Of the 21 non-minimal cases there were six of this kind, which is almost exactly what (5.64) would be expected by chance alone as calculated by the above method. A change at the mora level thus does not here seem to be influenced by similarity requirements at the submoraiic level.

3.6 Position in word

Analysis of the position of altered sounds in mutations must again take account of what is possible. Single-mora words do not occur among mutation source words. So initial moras may always be altered, as may the final (and non-initial) mora. Further, a medial mora can nearly always (in 86 of 93 cases) alter too. D does not choose at random from these options; 47 mutations are in initial moras (23 C, 8 V, 16 mora), 23 in medial (15:5:3) and 23 in final moras (17:3:3; $\chi^2[2] = 11.4$, $p < 0.01$). The tally of initial changes is higher than expected, and of medial and final changes lower than expected.

In the set of embeddings, we can also consider the position in the word of overlap between source word and pun; does the shorter form overlap with the initial portion of the longer form (as in 2f in Table 2), or with the final portion (as in 2e)? Initial overlaps are strongly favored, at 83:13 (eight final insertions and five final extractions), again a difference that departs highly significantly from chance ($z = 7.47$).

In the English lexicon, initial embeddings outweigh final embeddings (Luce, 1986; McQueen & Cutler, 1992), so dominance of initial embeddings might arise by chance. However, in Japanese this is not the case; the most common word lengths in the vocabulary (four- and five-mora words) have roughly as many final as initially embedded shorter words (the initial:final ratio is 1.18:1; Cutler, Otake, & Bruggeman, 2012). Such words account for 75% of the longer words in our embedding corpus (source words of extractions, puns in the case of insertions). Considering position only in this subset of 70 cases for which we know embedding statistics, the bias toward initial embedding is still significantly above chance (59 initial, 11 final: $\chi^2 = 25.39$).

We conclude that word-initial position is most likely to be altered in a mutation, and word-initial overlap is most likely to form the basis of an embedding.²

4 General discussion

This unusual opportunity to examine a substantial corpus of unrehearsed puns from an accomplished punster has provided insights into the processes of lexical selection involved in such word play.

D's long-running program has led him to build up a routine of facilitating light-hearted conversations. He asks factual questions of his interlocutors (*What are you studying? Where is your home?*) and the answers form the jumping-off point for his word play, whenever a play on words is what he chooses to create.

From the answers he receives, D selects a punnable source word. As the examples in Table 2 show, puns do not always occur to him immediately; repetition of the source word, as in (2g), (2i) or (2k), is common in our corpus. Our analyses suggest that his selection is meaning-driven; he seeks a source word and a punning transformation of it that is unexpected because it is amusingly different. This process results in a significant bias towards nouns as source words, because the answers nearly always contain a noun, nouns have contextually independent meaning, and the phonology of Japanese nouns yields candidate words with many close neighbors and often with homophones. Four source words recurred in our corpus, each time with a different pun outcome; for instance, *shinai* "downtown" was used for the homophone pun "bamboo sword" in one interaction, while in another interaction it was used for the homophone pun "not doing".

Although D thus works with whatever material his interlocutors happen to provide him with, our analyses motivate the conclusion that his punning is indeed motivated by what we have termed the MaxSP. Put differently, D has a very accurate conception of what constitutes a good pun.

The best pun is a homophone: the source-word sound pattern is completely preserved and only the meaning changes. D's preference for homophones where possible produces a corpus in which (a) the representation of homophones is roughly equivalent to their availability in the vocabulary, in which (b) the etymology of source words is skewed precisely in the case of homophones to over-represent that segment of the vocabulary that is most homophone-rich, and in which (c) mutated puns and many embeddings are very unlikely to have source words that are homophonic, although (d) homophone source words are highly likely to have possibly yielded alternative puns by processes of mutation or embedding.

How does a listener become conscious of a homophone? Firstly, consider the outline of the speech perception process presented in the introduction: speech input activates all words compatible with it, and these words compete for recognition. Words in memory have phonological representations, separate from their conceptual representations and, importantly, the competition process involves the former only (Norris, Cutler, McQueen, & Butterfield, 2006). Thus the meaning of a word arguably plays no role in the recognition process until evidence attests that the phonological form in question really has been heard. A listener who hears [bænd] will carry away just this phonological form from the competition process; only subsequent processing will enable the ongoing contextual construction to determine whether this was the word *band* or *banned*.

Given that recent research has shown that even the parameters of the competition process can be adjusted, such that phonologically mismatched words may not be immediately ruled out (Brouwer et al., 2012; McQueen & Huettig, 2012), we suggest that the parameters of contextual integration are similarly amenable to listener control; in other words, a listener searching for a homophone can adjust the contextual mismatch parameter so that, for instance, *banned* is not immediately ruled out in the context *I play in a* [bænd]. Inveterate punsters may even keep this contextual mismatch parameter quasi-permanently at a homophone-allowing setting. The structure of the spoken-word recognition process is thus in a sense conducive to such puns.

The same holds for puns with a part-whole relationship between source and pun. Here the experimentally attested modulation of mismatch in the competition process may be called into play. Hearing *plumbing* will activate *plum*, and hearing *cap* will partially activate *capture*; turning down the mismatch parameter will enable *plum* and *capture* to be carried through to later processing stages at which their conceptual representations can be consulted. In line with this suggestion, note that the shorter form most often overlapped with the initial portion of the longer form, both in the extractions and insertions in this corpus. Again, the punster requires no special operations to heighten alertness for word play; these parameter adjustments are available to cope with marginally difficult listening situations, such as casual speech (Brouwer et al., 2012; Mitterer, 2011) or mild background noise (McQueen & Huettig, 2012).

The spoken-word recognition system's word-play-friendly nature is actually a necessary consequence of the structure of vocabularies. All languages have a rich stock of words – tens or hundreds of thousands of them. However, languages prefer to have a relatively small stock of phonemes from which words are made up; the mean phoneme repertoire size is around 30 (Maddieson, 1984), and the modal size is 25 (Japanese has even fewer phonemes than modal; the English phoneme inventory, in contrast, verges on profligate abundance, especially in the British variety with 46 phonemes). With only a handful of phonemes even in the most well-endowed inventories, then, and so many words to construct from them, the inevitable result is that words are likely to resemble one another and overlap with one another. The smaller the phoneme inventory, the longer words are likely to be, but the greater the number of embedded forms (Cutler, Norris, & Sebastián-Gallés, 2004). Even though the type and location of embedding differs with variation in phonological structure (Cutler et al., 2012), all languages have a highly embedded stock of words. Homophony,

too, is rife; although it may seem most likely when phoneme inventories are small, no language actually resists homophony, and the processes of new word creation tend, across languages, to assign new meanings to old forms in preference to creating new forms. Ambiguity, such as that expressed in re-use of homophonous forms, in fact serves the interest of language users by reducing effort (Piantadosi, Tily, & Gibson, 2012; Zipf, 1949). The fact that it makes word play easy is a happy side-effect.

What, then, if the dedicated punster finds no homophone and no part–whole relationship readily available in an incoming utterance? The remaining option is to alter the source word's form to make another word similar to it. Mutation, our analyses suggested, is D's least favored option (together the two source-preserving pun types significantly outnumbered mutations; mutation source words largely had no homophones, while homophones and most embeddings could easily have become mutations instead). However, when a mutation is undertaken, there are clear signs that Maximal Source Preservation remains in force.

Most importantly, alterations to the source word are minimal, by preference confined to a single phoneme, which by preference should overlap with the replaced phoneme by two broad features; for instance, *peas* would be more likely to be replaced by *bees* or *keys* than by *knees* or *cheese*. This finding is consistent with our earlier argument, based on experimental evidence and on analyses of varying types of puns (Cutler & Otake, 2002; Otake, 2010; Otake & Cutler, 2001), that spoken-word processing in Japanese is continuous and not mediated by supraphonemic levels of representation, such as the mora. Similarity between Japanese words is evaluated in the same way as similarity between words in any other language. Consistent with this, studies of phonetic confusions in Japanese (Saito, 1961) and in English (e.g., Miller & Nicely, 1955) produce parallel patterns (e.g., greater confusability of place of articulation, lesser confusability of manner of articulation), and similarity ratings for pairs of simple CV syllables are highly consistent across Japanese and English, as well as Italian and Serbo-Croatian (Takefuta, Guberina, Pizzamiglio, & Black, 1986).

Two aspects of our analyses of similarity require further discussion, however. One is the failure of the pitch accent pattern of words to play a definitive role as a criterion of source preservation. This is not to say that accent pattern varied randomly; in contrast, accent pattern was largely preserved in homophones and mutations, where length of source and of pun was constant. However, where length of source and pun varied, there was no sign of a preference for accent preservation (in line with the strong correlations between accent pattern and word length in Japanese; Sugito, 1995).

It may seem that this finding suggests a different picture than that arising from the role of stress in English puns (where puns that shift stress are disfavored; Lagerquist, 1980). However, this may not be the case. For listeners to English, the segmental consequences of stress changes outweigh the suprasegmental realizations; listeners find cross-splicings acceptable between syllables with the same vowel but different stress levels (e.g., the initial syllables of *autumn* and *automata*; Fear, Cutler, & Butterfield, 1995), and mis-stressing that does not change a vowel does not significantly inhibit word recognition (Bond & Small, 1983; Slowiak, 1990; Small, Simon, & Goldberg, 1988). In contrast, the best way to disrupt word identification in English is to alter a vowel in a stressed syllable (Bond, 1981). These findings suggest that word-form similarity is measured on a purely segmental scale in English, with suprasegmental realization contributing relatively little. Because English stress variation so often has a segmental reflection, segmental processing accomplishes nearly all removal of spurious competition in spoken-word recognition, leaving little role for suprasegmental cues (Cutler et al., 2004). In Japanese, accent placement indeed informs listeners' word recognition (Cutler & Otake, 1999; Sekiguchi & Nakajima, 1999). However, the

similarity analyses for these Japanese mutating puns suggest that the scale on which inter-word similarity is computed is comparable for each language.

Another important finding in our analyses of similarity between source and mutation pun concerned the position where such minimal segmental changes were made. There was a very strong tendency for the change to be made word-initially, thus preserving overlap between the latter part of the source and pun words at the expense of the onset. The counts that we made of the options for mutations that would produce another word indicated that substitution possibilities were not clustered at word onset in the way the observed mutations were (e.g., for example [2g], the source word *rikei* could have yielded 13 other words and replacement of any of the phonemes could have produced a word); that is, this finding cannot be explained in terms of the structure of the vocabulary, but is a strategic preference.

Further, this positional effect for segment mutation is not immediately susceptible to explanation in terms of the nature of the spoken-word recognition process, as was the case, for example, with the homophony patterns or indeed with the positional effect in embedding patterns. Consider that the effects of a mismatch in the competition process will be felt as soon as the mismatching phoneme arrives; the earlier arrival of mismatch in word onset than later in the word is the source of the asymmetry between offset and onset competitors described in the introduction. The kind of competitor that stays around longest in normal spoken-word recognition – given the input *candle* for example, the *candy* that persists well after the *handle* has stopped competing – is **not** the kind of competitor that is chosen as a pun. Instead, on our corpus results, *handle* would be more likely than *candy* to be selected as a mutation of *candle*.

This pattern may arise by speaker choice, in that explicit search for a mutation is initiated from the word-initial mora. The corpus of 93 mutations does not reveal any distinct signs of this, such as perhaps ordered preference for particular moras. In an equivalent English corpus, one might, for example, test for alphabetical order preferences in rhyme choices. Japanese does not have rhyming traditions, however, and explicit order preferences may differ across individuals.⁶ One option, supported by the strong tendency to featural overlap in the mutation pun set, would be articulatorily based generation⁷ (consider here that D quite often repeats his chosen source word before punning on it in a subsequent utterance). Note, however, that investigations of the basis for explicit phonological similarity judgments have revealed that listeners rate onset alterations with rhyme preservation (*sash-fash*) as more highly similar than coda alterations with onset plus vowel preservation (*sash-saf*; Hahn & Bailey, 2005). Thus, once again, D's choice is actually one that accords well with listener similarity preferences.

In short, this study has allowed us to observe the processes that underlie successful punning. At the most general level, two conclusions are clearly warranted: firstly, the normal processes of spoken-word recognition essentially facilitate word play, in that they offer numerous opportunities for listeners to capitalize on intermediate processing stages to find similar words without departing from standard processing procedures and, secondly, accomplished punsters target those puns that will work well, by applying a MaxSP that delivers the same result as experimentally tested indices of inter-word similarity (and as offline collections of puns chosen because they work well). Given that we have based this study on data from a single highly accomplished individual, only empirical exploration can ascertain whether the same patterns will hold in data collected from other speakers and in other languages. Our limited exploration of the interlocutor puns in our corpus (note 2), the similarity of the patterns noted here to patterns observed in published pun collections and internet databases and, most importantly, the naturalness of D's punning processes and the close relationship of these processes to the known mechanisms of speech perception, all suggest that D is exceptional not in the manner in which he puns, but only in the fluency and mastery of his punning.

Nonetheless, we would certainly hope that this study prompts further empirical investigation; although, as noted in the introduction, the analysis of natural speech phenomena accounts for only a small proportion of psycholinguistic research, it can yield rich lodes of valuable data.

Acknowledgement

We are grateful to Kazuhiko Kakehi for assistance with obtaining phonetic confusion data for Japanese.

Funding

This work was supported by the NWO-SPINOZA project “Native and nonnative listening” (A. Cutler).

Notes

1. In recent years some of the recipients of D’s attention have taken the opportunity of making amateur videos of his visit to their workplace, and readers who understand Japanese may enjoy several such videos on YouTube (e.g., a visit to a printing company at <http://www.youtube.com/watch?v=gsK1tyOEhE&feature>).
2. We analyzed the 62 puns collected from other speakers. Very much the same patterns appeared as in D’s puns (there were 22 homophones, 29 embeddings and 18 mutations; all but three source words and all but six pun words were nouns; 10 of the mutations were word-initial and 14 involved a single phoneme).
3. Non-noun source words showed no deviant patterns, but it is worth noting that all fell in the embedding category and yielded noun puns – e.g., *sabasaba* “refreshed-looking” yielded *saba* “mackerel”, *kawara-nai* “unchanged” was reparsed as *kawara nai* “no tiles”.
4. Indeed, the spontaneous interactions consisted mostly of short utterances.
5. The embedding category included one full reparse, the *kawaranai* case noted above.
6. Acoustic explorations of the recorded conversations also produced no systematic patterns, either by pun type or by characteristics of mutation puns.
7. Kawahara and Shinohara (2009), in their analysis of internet puns, argued for a perceptual rather than articulatory basis for similarity. However, they applied abstract phonological criteria (e.g., comparing null onsets to consonantal onsets) that reduce the relevance of articulatory features.

References

- Al-Kisai, A. (1915). Speech errors of the populace. (In Arabic; Ninth-century MS edited by C. Brockelmann.) *Zeitschrift für Assyriologie und verwandte Gebiete*, 13, 29–46.
- Bond, Z. S. (1981). Listening to elliptic speech: Pay attention to stressed vowels. *Journal of Phonetics*, 9, 89–96.
- Bond, Z. S., & Small, L. H. (1983). Voicing, vowel, and stress mispronunciations in continuous speech. *Perception & Psychophysics*, 34, 470–474.
- Brouwer, S., Mitterer, H., & Huettig, F. (2012). Speech reductions change the dynamics of competition during spoken word recognition. *Language and Cognitive Processes*, 27, 539–571.
- Cutler, A. (1981). The reliability of speech error evidence. *Linguistics*, 19, 561–582.
- Cutler, A. (2012). *Native listening: Language experience and the recognition of spoken words*. Cambridge, MA: MIT Press.
- Cutler, A., Norris, D., & Sebastián-Gallés, N. (2004). Phonemic repertoire and similarity within the vocabulary. In S. H. Kin, & M. Jin Bae (Eds.), *Proceedings of the 8th International Conference on Spoken Language Processing*, Jeju Island, Korea (Vol. 1, pp. 65–68). Seoul: Sunjin Printing Co.
- Cutler, A., & Otake, T. (1994). Mora or phoneme? Further evidence for language-specific listening. *Journal of Memory and Language*, 33, 824–844.
- Cutler, A., & Otake, T. (1999). Pitch accent in spoken-word recognition in Japanese. *Journal of the Acoustical Society of America*, 105, 1877–1888.

- Cutler, A., & Otake, T. (2002). Rhythmic categories in spoken-word recognition. *Journal of Memory and Language*, 46, 296–322.
- Cutler, A., Otake, T., & Bruggeman, L. (2012). Phonologically determined asymmetries in vocabulary structure across languages. *Journal of the Acoustical Society of America*, 132, EL155–160.
- Fear, B. D., Cutler, A., & Butterfield, S. (1995). The strong/weak syllable distinction in English. *Journal of the Acoustical Society of America*, 97, 1893–1904.
- Ferber, R. (1991). Slip of the tongue or slip of the ear? On the perception and transcription of naturalistic slips of the tongue. *Journal of Psycholinguistic Research*, 20, 105–122.
- Ferber, R. (1993). *Wie valide sind Versprechersammlungen?* Bern, Switzerland: Peter Lang.
- Hahn, U., & Bailey, T. M. (2005). What makes words sound similar? *Cognition*, 97, 227–267.
- Katada, F. (1990). On representation of moras: Evidence from a language game. *Linguistic Inquiry*, 21, 641–646.
- Kawahara, S., & Shinohara, K. (2009). The role of psychoacoustic similarity in Japanese puns: A corpus study. *Journal of Linguistics*, 45, 111–138.
- Kindaichi, K., Saeki, U., Oishi, H., & Nomura, M. (2002). *Shinsen Kokugo Jiten* (“Newly selected Japanese dictionary”), 8th edition. Tokyo, Japan: Shogakukan.
- Kubozono, H. (1989). The mora and syllable structure in Japanese: Evidence from speech errors. *Language and Speech*, 32, 249–278.
- Kubozono, H. (1990). The role of the mora in speech production of Japanese. *Proceedings of the International Conference on Spoken Language Processing* (Vol. 1, pp. 501–504). Kobe, Japan: Acoustical Society of Japan.
- Kubozono, H. (1998). A syllable-based generalization of Japanese accentuation. In R. H. Mannell & J. Robert-Ribes (Eds.), *Proceedings of the 5th International Conference on Spoken Language Processing, Sydney*. Canberra, Australia: Australian Speech Science and Technology Association (ASSTA).
- Lagerquist, L. M. (1980). Linguistic evidence from paranomasia. In *Papers from the 16th Regional Meeting, Chicago Linguistic Society* (pp. 185–191). Chicago, IL: Chicago Linguistic Society.
- Luce, P. A. (1986). A computational analysis of uniqueness points in auditory word recognition. *Perception & Psychophysics*, 39, 155–158.
- Maddieson, I. (1984). *Patterns of sounds*. Cambridge, UK: Cambridge University Press.
- McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18, 1–86.
- McQueen, J. M. (2007). Eight questions about spoken-word recognition. In M. G. Gaskell (Ed.), *The Oxford handbook of psycholinguistics* (pp. 37–53). Oxford, UK: Oxford University Press.
- McQueen, J. M., & Cutler, A. (1992). Words within words: Lexical statistics and lexical access. In J. J. Ohala, T. M. Nearey, B. L. Derwing, M. M. Hodge, & G. E. Wiebe (Eds.), *Proceedings of the International Conference on Spoken Language Processing '92, Banff* (Vol. 1, pp. 221–224). Edmonton, Canada: University of Alberta.
- McQueen, J. M., & Huettig, F. (2012). Changing only the probability that spoken words will be distorted changes how they are recognized. *Journal of the Acoustical Society of America*, 131, 509–517.
- McQueen, J. M., Norris, D., & Cutler, A. (1994). Competition in spoken word recognition: Spotting words in other words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 621–638.
- Meringer, R., & Mayer, C. (1895). *Versprechen und Verlesen: Eine psychologisch-linguistische Studie*. Stuttgart, Germany: Göschen.
- Miller, G. A., & Nicely, P. E. (1955). An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America*, 27, 338–352.
- Mitterer, H. (2011). Recognizing reduced forms: Different processing mechanisms for similar reductions. *Journal of Phonetics*, 39, 298–303.

- Nakano, H. (1989). Kookoo Kyookasho no Dooongo (Homonym words in high school text books). In *Report of National Institute for Japanese Language 99: Kookoo Chuugaku Kyookasho no Goichoosa* ("Studies on the vocabulary of high and middle school textbooks") (pp. 77–131). Tokyo, Japan: Shuei Shuppan.
- National Institute for Japanese Language (2006). *Gendai zasshi 200 manji gengo choosa goihyoo* CD ver. 1.0 ('Linguistic survey of two million characters in contemporary Japanese magazines'). Tachikawa, Japan: NIJL.
- Norris, D., Cutler, A., McQueen, J. M., & Butterfield, S. (2006). Phonological and conceptual activation in speech comprehension. *Cognitive Psychology*, 53, 146–193.
- Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review*, 115, 357–395.
- Otake, T. (2010). *Dajare* is more flexible than puns: Evidence from word play in Japanese. *Journal of the Phonetic Society of Japan*, 14, 76–85.
- Otake, T., & Cutler, A. (2001). Recognition of (almost) spoken words: Evidence from word play in Japanese. In P. Dalsgaard, B. Lindberg, H. Benner & Z.-H. Tan (Eds.), *Proceedings of EUROSPEECH 2001*, Aalborg (pp. 465–468). Aalborg, Denmark: Kommunik Grafiske Løsninger.
- Otake, T., Hatano, G., Cutler, A., & Mehler, J. (1993). Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language*, 32, 358–378.
- Piantadosi, S. T., Tily, H., & Gibson, E. (2012). The communicative function of ambiguity in language. *Cognition*, 122, 280–291.
- Saito, S. (1961). Confusion matrices of Japanese speech sounds. *Electrical Communication Laboratory Technical Journal*, 6, 1–159.
- Sekiguchi, T., & Nakajima, Y. (1999). The use of lexical prosody for lexical access of the Japanese language. *Journal of Psycholinguistic Research*, 28, 439–454.
- Sherzer, J. (1970). Talking backwards in Cuna: The sociological reality of phonological descriptions. *Southwestern Journal of Anthropology*, 26, 343–353.
- Siegel, S. (1956). *Nonparametric statistics for the behavioral sciences*. New York, NY: McGraw-Hill.
- Slowiaczek, L. M. (1990). Effects of lexical stress in auditory word recognition. *Language and Speech*, 33, 47–68.
- Small, L. H., Simon, S. D., & Goldberg, J. S. (1988). Lexical stress and lexical access: Homographs versus nonhomographs. *Perception & Psychophysics*, 44, 272–280.
- Stemberger, J. P. (1985). *The lexicon in a model of language production*. New York, NY: Garland.
- Sugito, M. (1995) *Akusento onsei jiten (Accent pronunciation dictionary) CD ROM*. Tokyo: Maruzen.
- Takefuta, Y., Guberina, P., Pizzamiglio, L., & Black, J. W. (1986). Cross-lingual measurements of interconsonantal differences. *Journal of Psycholinguistic Research*, 15, 189–507.
- Tokuhiki, Y. (2005) Nihongo dooongo no akusento to onsetsu koozoo ("Accent and syllable structure in the Japanese homonym"). *Hakuoo Daigaku Ronshuu* ("Hakuoo University Bulletin"), 19, 45–62.
- Vitevitch, M. S. (2002). Influence of onset density on spoken-word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 28, 270–278.
- Zipf, G. K. (1949). *Human behavior and the principle of least effort*. Reading, MA: Addison-Wesley.