Loanword formation: a neural network approach

Alan D. Blair
blair@cs.uq.edu.au
John Ingram
jingram@lingua.cltr.uq.edu.au

Department of Computer Science and Electrical Engineering
Department of English
University of Queensland, 4072, Australia

Abstract

Loanword phonology seeks to model the process by which foreign words are ‘nativised’ or incorporated into the phonological system of the ‘borrowing’ language. We can conceive of this as a parsing of the phonetic *input* provided by the foreign word forms, in accordance with phonological *output* constraints of the borrowing language. Following Silverman (1992) we conceive loanword formation as fundamentally a *two-stage* process: the first of which yields a parsing of the phonetic input into segmentally organised phonetic feature bundles, interpretable as segmental targets in the borrowing language. In the second stage of processing, these segmental targets are parsed into phonological structures (syllables, mora, feet...etc) compatible with the word-prosody of the borrowing language.

Japanese borrowings from English provide a good test-bed for models of loanword formation, because examples are abundant and, while the segmental mapping from English to Japanese is relatively straightforward, their respective word-level prosodies are strikingly different, providing ample opportunity to observe prosodic re-structuring in loanword formation.

The theoretical aim of this study is to investigate the learning mechanisms required for phonological parsing in loanword formation. Phonologists have typically sought to construct ‘symbolic’ or ‘analytical’ parsing algorithms for this task, either with or without reference to a framework of learnability theory. This approach came under strong challenge from ‘empirical’ Connectionist models of language processing in the 1980's. The debate between these competing paradigms, or the search for some suitable ‘hybrid’, continues. It may be argued that loanword formation provides a more restrictive, and hence better controlled, environment for studying parsing mechanisms than other natural language processing tasks, which involve a host of lexical, morphological, syntactic or pragmatic influences.

In this paper, we report the results obtained from a feed-forward neural network, trained on an 1100 word corpus of American English loanwords `borrowed’ into Japanese, post-WWII. We make preliminary comparisons with an analytical constraint-based approach to modeling loanword formation.

An intended application for the NN parser, was to devise an English-Japanese proper or place-name translator, which would map an English phoneme sequence into Katakana (e.g.: *Brisbane* /brɪzbən/ => ﾌﾞ ﾘｽﾞ ﾍﾞ ﾝ).

Introduction

The Japanese language has borrowed thousands of words from English, particularly since World War II, under the overwhelming economic and cultural influence of the United States. This massive borrowing over a comparatively short period of time provides a unique window on processes of loanword formation. Borrowed words may reflect varying degrees of nativisation to the phonological patterns of the borrowing language.

While the segments and phonetic features of English words tend to be remarkably well preserved by the process of loanword formation, the resulting Japanese word forms are so completely transformed in terms of their prosodic structure that English listeners almost invariably fail to recognize their English sources, when loanwords are presented to them as isolated words carefully spoken by a native speaker of Japanese (Ingram, 1998). The main factor underlying poor recognition of the English source words appears to lie in the extensive resyllabification, involving vowel epenthesis, which is required to parse the segmental input into Japanese prosodic frames. Some examples are given below:

orthographic	phonemic	Japanese
Olympic	əlimpik	oriNpikku
truck	trak	torakku
cut	kat	katto
cud	kad	kado
cart	kart	kaato
cat	kæt	k^jatto

Japanese syllable structure is predominantly CV, lacking complex onsets and codas. To maintain faithful representation of the segment structure of the English source words, extensive use is made of the epenthetic vowels (chiefly, /u/, /o/ and /I/). The temporal structure of the English source word is converted to Japanese moraic timing. English tense (long) vowels and diphthongs are usually treated as two-mora units, whereas lax (short) vowels are assigned to a single mora. (Stress pattern in the English source word plays a moderating role. Tense vowels or diphthongs in unstressed syllables may emerge as one-mora vowels, as in *Olympic* [əlɪmpɪk].)

English voiceless obstruents following a lax vowel are almost always treated as geminate (two mora) stops. But voiced obstruents usually do not geminate in this environment in loanword formation. Also, if the preceding vowel is tense (long) in the English source word, gemination of voiceless obstruents does not occur in the loanword. The reasons for these timing changes need not concern us here.

The segmental mapping from the English source word to segments in the Japanese loans is basically one-for-one, without any feature exchange between adjacent segments. There are some palatalized consonants in Japanese borrowings which appear to 'pick up' the palatalization feature from the following vowel (see, *cat* above). However, this feature-swapping between adjacent segments occurs only under quite restrictive conditions. Note that Japanese has no sound which corresponds to the low *front* English /æ/. We hypothesise that Japanese listeners respond to the 'palatal' quality of English /æ/, by relocating this feature onto the preceding consonant, thereby regularizing the foreign phonetic contrast to Japanese phonemics.

The examples given above illustrate the major phonological transformations required to parse English source words into English loanwords in Japanese, within the limits of a phonemic transcription. Implicit in these transformations, though they are not explicitly captured by the phonemic representation, are the contrasting prosodic structures, such as syllable structure constraints, which motivate the transformations.

Modeling loanword formation

In an 'analytic' account of loanword formation, these prosodic constraints would be explicitly represented in the form of re-write rules, filters or constraints on well-formedness for particular aspects of prosodic structure. The output of parsing would take the form of assignment of an explicit prosodic phrase marker. By contrast, in a NN based account of loanword formation, the details of prosodic feature assignment are not explicitly represented, but are typically regarded as *emergent* features, contained in the weight-states of the trained network. The NN demonstrates acquisition of the prosodic constraints on word formation by being able to transform phonetic segments and features of input source words into well-formed phonemic sequences in the borrowing language. (The further step of transforming Japanese phonemic representations into Katakana script is trivial.)

In a complete model of loanword formation, it would of course be necessary to specify more precisely how the stage I extraction of phonetic segments and features from the speech signal is achieved and what perceptual filtering takes place on the segmental phonetic properties of foreign words as processed through the ear of the native listener. These initial auditory representations are not, strictly speaking, sequences of *phonemic* segments or features. Rather, they are likely to be evanescent and somewhat fragmentary, made up of phoneme-like segments and 'foreign' phonetic features of sufficient auditory saliency that they cannot be ignored by the perceptual mechanism.

However, for purposes of modeling stage II processes of loanword formation, whether an analytical or a statistical NN model is adopted, an initial segmental parsing of the input word into sequences of phoneme-like feature bundles was assumed (see Appendix I for the segments and features used).

The Database

We compiled a database of 1100 words from a dictionary of neologisms borrowed mainly from American usage in the post-war era (Bailey, 1962).

The English phonemic transcriptions of these words were obtained from the Carnegie Mellon Pronouncing Dictionary (ftp://ftp.cs.cmu.edu/project/fgdata/dict/) which generally reflects American rather than British pronunciation.

Network architecture

We use a two-layer feed-forward neural network with 65 inputs, 20 hidden units and 53 outputs. A featural representation is employed for the (English) input and a phonemic representation for the (Japanese) output. This architecture was inspired by NETtalk (Sejnowski & Rosenberg, 1987) although the task performed is, in a sense, the reverse of that performed by NETtalk, since the latter used an orthographic representation for its input and a featural representation for its output.

If the aim of the exercise were to model the way humans learn the task of loanword formation, it would be more appropriate to train the network on some task involving only the target language, prior to testing it on input from the source language. In that scenario, the training task would presumably involve deriving the correct surface form from some hypothesised underlying structure. However, many choices and assumptions would need to be made in order to devise an appropriate framework and representation for these underlying structures. By avoiding such choices, our approach has the advantage of allowing phonological constraints to be studied in a more canonical context. But it means that the network is really performing a composite task, since it can make use of the constraints and statistics of the source language as well as those of the borrowing language.

The 65 inputs are divided into 5 groups of 13, which encode the phonological features of the current phoneme, the two preceding phonemes and the two following phonemes (13 features × 5 phonemes = 65 inputs in all).

The featural input representation has several advantages over a phonemic one:

(1) it reduces the number of inputs,
(2) features often influence the form of loanwords in a systematic way,
(3) the same word is often rendered differently in different dialects (for example, British vs. American English) and the featural representation is less sensitive to this than a phonemic one would be.

Input and output phonemes do not always correspond on a 1-to-1 basis. In some cases a phoneme may be deleted, or it may have a consonant and/or a vowel appended to it. In order to allow for these possibilities, we divide the outputs of our network into three groups. The first group has one output for each possible phoneme (consonant or vowel); the second group has one output for each possible consonant; the third group has one output for each possible vowel. Each group has one additional output representing the `empty' phoneme "_". Since there are 20 consonants and 5 vowels in Japanese, the total number of outputs is 26+21+6=53.

For example, consider the English word *cat* which has the phonemic representation /kæt/ and becomes /kjatto/ in Japanese. The network views this as three separate training items:

_ _ k æ t => k j _
_ k æ t _ => a _ _
k æ t _ _ => t t o

This means that the network, when presented with the features encoding the input /_ _kæt/, should be trained to produce an activation of 1.0 for the /k/ output of the first group, the /j/ output of the second group and the /_/ (empty) output of the third group (and an activation of 0.0 for the other 50 outputs).

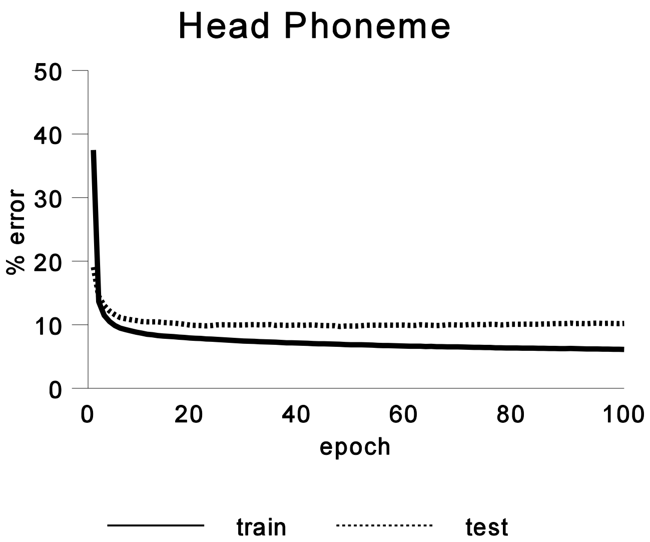
When it comes to the testing phase, within each group the output with the largest activation is selected, and this determines the three-phoneme sequence chosen by the network to correspond with the salient input phoneme.

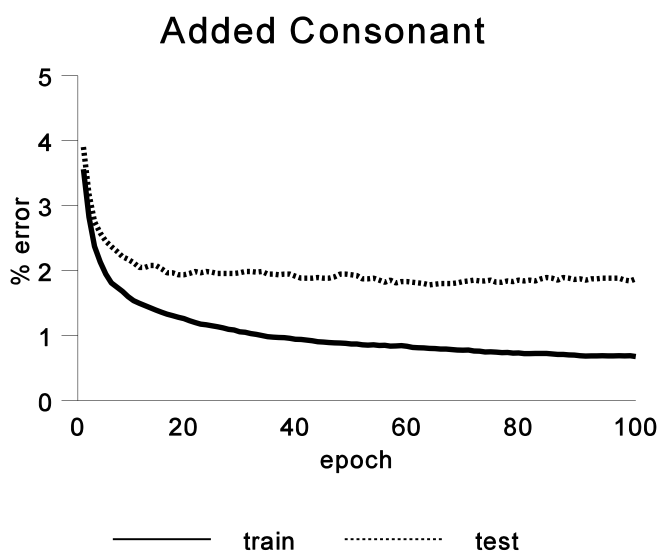
The networks were trained by back-propagation (Rumelhart et al., 1986) for 100 epochs, with a learning rate of 0.01 and a momentum of 0.9. The cross-entropy minimization criterion was used.

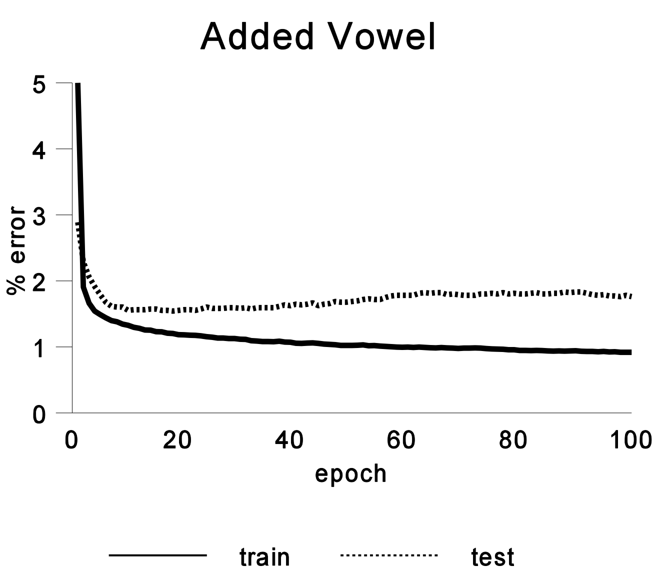
Results

Each of 11 networks was trained on 1000 words from the database, and tested on the other 100 words. Each word occurred in the test set of exactly one network.

The 1100 words in the database had an average of 8.8 phonemes per word, making a total of 9658 input phonemes. Each of these input phonemes can produce output consisting







of a head phoneme (group 1 outputs), plus an optional added consonant (group 2) and/or added vowel (group 3).

Figures 1-3 show the percentage of errors on the training and test sets for each of these three groups. In our data set, the head phoneme was nonempty 97% of the time, while the added consonant and added vowel were nonempty only 4% and 17% of the time, respectively, so the network error is much smaller for the latter two groups. After 30 epochs the training and test errors, respectively, reach a level of 7.4% (resp. 9.9%) for the head phoneme, 1.1% (resp. 2%) for the added consonant, and 1.1% (resp. 1.6%) for the added vowel. After this, the training error continues to fall while the test error levels off. (Note: the test error was computed at the end of each epoch, while the training error was computed during the epoch. Therefore the training error may exceed the test error in the first few epochs.)

Error Analysis

An analysis was undertaken of all 'errors': cases where there was a discrepancy between the romanji (Japanese Romanized) representation of a loan word and the phonemic representation assigned by the fully trained network, when the item in question was not included in the training set. These are summarised in Table 1. The 'error' categories are not necessarily mutually exclusive; nor do they necessarily indicate an error on the part of the network, but simply a discrepancy between the dictionary-based phonemicization (Romanji transcription) and that assigned by the network.

Discrepancies of Schwa vowel colouring:

The most common 'error' or discrepancy between the network-assigned Japanese phonemicization and the Romanji dictionary entry for loanwords (27% of cases) involved unstressed vowels in English source words.

Table 1. Primary "error" count

Error or discrepancy	Incidence	
Schwa vowel colouring	148	27%
Vowel length discrepancy	107	19%
Obstruent gemination	45	11%
Vowel epenthesis	39	7%
Dipthongs /ou/ and /ei/	23	5%
Spelling pronunciation	20	4%
Post vocalic /r/	20	4%
Vowel quality, unexplained	20	4%
Palatalization of /t/ to /ʧ/	18	3%
Back vowel phonemicization	10	2%
Epenthetic vowel quality	10	2%
Alternation: /ʤ/ and /z/	8	2%
Silent /g/ in *ng*	7	1%
Alternation: /w/ and /u/	7	1%
Alternation: /s/ and /ʃ/	7	1%
Alternation: /h/ and /f/	6	1%
/y/ epenthesis: breaking	5	1%
Others	8	2%
TOTAL	489	

Such vowels would normally be transcribed with a schwa [ə] in English phonemic representations. Japanese has no equivalent to schwa and Romanji transcriptions of such vowels are guided by English spelling in the selection of an appropriate symbol. Because the network had no access to the orthographic representations of English source words, these discrepancies were frequent.

Discrepancies of vowel length:

The second most frequent error involved discrepancies of vowel length. English tense vowels and diphthongs should be perceived as long (two mora) vowels in Japanese. However, stress and position in the word may act as moderating influences. In primary stressed position English vowels are lengthened, while reduced vowels in unstressed syllables may be very short. Therefore, tense English vowels in unstressed position may not be perceived as long (bimoraic) by Japanese listeners.

An analysis of discrepancies between the network predictions and the romanji assigned vowel durations (Table 2) revealed that only 13% of cases (example [1]) could be accounted for by shortening of a tense vowel in an unstressed syllable.

Table 2. Discrepancies of Vowel Length*

Romanji form of loanword / Network assigned Romanji	English orthography / English phonemic	Incidence	
1. a_f_ut_a_n_u_u_N	afternoon		
a_f_ut_**a**_**a**_n_u_u_N	æftərnuun	14	13%
2. a_k_ut_i_b_i_t_i_i_z_u	activities		
a_k_uC_i_b_i_t_**i**__z_u	æktivətiiz	42	39%
3. a_n_a_r_o_J_i_i_	analogy		
a_n_a_r_o_J_**i**__	ənæləʤii	21	20%
4. a_d_ob_a_N_t_**e**_**e**J_i	advantage	3	03%
a_ddob_e_N_C_i_JJi_	ədvæntiʤ		
5. a_N_C_o___b_**e**__	anchovy	14	13%
a_N_C_o_o_b_i_i_	æntʃouvii		
6. C_e_kkuo__f_u	checkoff		
C_e_k__o_o_f_u	ʧekoof	8	07%
7. a_u_t_oo_b_uf_a_SS_o_N_	out-of-fashion		
a**u**_**u**t_oa_b_uf_a_SS_o_N_	autavfæSən	3	03%

*Source of discrepancy indicated in **bold**.

In the majority of cases (59%), the

discrepancy between the romanji and the network assigned vowel length was caused by the network shortening phonemically long/tense vowels in final position (example 2[2.]) or elsewhere in the word (2[3]). We are presently unable to account for this behaviour of the network. In 13% of cases (2[5.]) vowel length discrepancies could be sourced to irregular romanization on the part of the compilers of the Japanese dictionary, or to errors of vowel length phonemicization in the the American dictionary (7%, 2[6]). The influence of English spelling could be clearly seen in the romanji forms in 3% of cases: e.g., *advant**a**ge* => adobaNt**ee**Ji because *age* => eeJI, though the pronunciation [ədvæntɪʤ] indicates the vowel is short/lax. In a small proportion of cases (3%, 2[7]) the vowel length discrepancy appeared to be attributable to the network's allowance of three-mora vowel sequences within a single syllable (super-heavy syllables), not sanctioned by Japanese syllable structure.

Gemination of obstruents:

The basic rule of gemination for English loanwords is that voiceless obstruents geminate following short vowels in stressed syllables. Voiced stops also geminate irregularly in this environment. The most common error of gemination involved the network failing to geminate a voiceless obstruent in the expected environment (Table 3, example [1]). But the network also inappropriately produced geminates following an unstressed vowel (3[2]), though not consistently. We observed that gemination did not occur in romanji forms derived from a consonant cluster in the English word (e.g.: the /k/ in *vector* does not geminate, becoming Romanji /be**k**utoru/). However, this constraint was not respected by the network (3[3]). The network occasionally produced geminate consonants linked to its propensity to shorten vowels (3[4]). Gemination of voiced obstruents was irregular in the Romanized forms (3[5,7]) and consequently in the network output as well (3[4,5]). Nor was the gemination of voiceless obstruents entirely regular in the romanji forms (3[9]).

Table 3 Errors of Consonant Gemination*

_desired_output_ _actual__output_	English-orthography English-phonemes	Incidence
1. a_N_r_a__kk_i_i_ a_N_r_a_**k**__i_i_	unlucky-net ənlakiinet	18
2. a_n_e_k_ud_o_o_t_o a_n_e_**kk**ud_o___t_o	anecdote ænəkdout	12
3. b_e_k_ut_o_r_u b_e_**kk**ut_o_a__	vector vektər	10
4. f_i_i_d_ob_a_kku f_i__**dd**ob_a_kku	feedback fiidbæk	3
5. b_o_**bb**us_ur__e__e__ b_o__b_us_ur__e__e__	bob-sleigh bobslei	3
6. a_d_ob__a__N__s_u a_**dd**ob__a__N__s_u	advance ədvæns	1
7. a__d_or__I__b_u a__**dd**or__I__b_u	ad-lib ædlib	1
8. a_i_ky_a__CC_a__a__ a_i_**kk**_a__CC_a__a__	eye-catcher aikæʧər	2
9. f_e_**t**_i_S_i_z_u_m_u f_e_tt_i_S_i_z_u_m_u	fetishism fetiʃizəm	4

* The inferred site of error is indicated in **bold**.

Vowel epenthesis:

Errors of vowel epenthesis are of particular interest in assessing the network's capacity to adapt to Japanese syllable structure. Epenthetic vowels are very frequent in English loan words (averaging 1.28 per word in romanji forms in the current data set, or slightly over 1400 occurrences). However, only 39 discrepancies of vowel epenthesis were observed (an error rate of 3%).

Furthermore, the largest sub-category of epenthesis errors were related to word boundaries in compound forms, (see Table 4).

Table 4. Epenthesis at word boundaries*

desired output actual output_	English-orthography English-phonemes	Incidence
a_t_o_m_i_kk **u**e_i_J_i a_t_o_m_i_kk_e_e_J_i	atomic-age ətomikeiʤ	
d_or_a_i_b_**u**i_N_ d_or_a_i_b_i_N_	drive-in draivin	11

*Epenthesis associated with boundary in **bold**

The network, having no information about word boundaries, could not predict epenthetic vowels at word endings in compound forms. However, such boundaries are inherently problematic in loan word formation, as they may or may not be apparent to speakers of the borrowing language.

The quality of the epenthetic vowel ([I], [u], [a], [e], [o]) was also well predicted by the network, with only 10 discrepancies (<1%); in all cases caused by irregular romanization, some of which were clearly related to English spelling.

Dipthongs /ou/ and /eɪ/:
The diphthongs /ou/ and /eɪ/ are usually represented as long vowels ‘oo’ and ‘ee’ in romanji, but not consistently. In approximately 40% of cases, English /ei/ was rendered as Romanji ‘ei’ and in 32% of cases English /ou/ converted to the short vowel ‘o’ in Romanji. These irregularities in the romanji representation of the diphthongs /ei/ and /ou/ are the probable cause of the occasional discrepancies with the network predictions.

In summary, the network’s performance in predicting the phonological forms of loan words in Japanese, was on the whole, very good, except for the features of vowel quality in reduced syllables of English words, and in the prediction of vowel length and consonant gemination in romanji forms. Except in the case of English vowels in reduced syllables, spelling was found to play a subordinate role to the phonological features of words in the source language. Some improvement in the prediction of vowel length, and substantial improvement in the prediction of consonant gemination may be expected by providing the network access to the locus of primary stress in English words. We are currently investigating this.

Discussion

We confine discussion to observations on the main practical and theortical objectives of this on-going study. With respect to the goal of devising an English-Japanese proper or place-name translator, which will convert English phonemic representations to romanji or kana forms, we find the results encouraging. With the inclusion of primary stress, and access to orthographic representations for the prediction of vowel quality in reduced syllables, it should be possible to obtain near optimal performance for the prediction of romanji or kana forms, given a degree of indeterminism that is present in loan word dictionary entries. Precisely how English orthographic information can be incorporated as required is a problem that we have not yet addressed. For the data set of our study (post-war borrowings of general lexical items), it is clear that English spelling plays a strictly subordinate role to the phonetic form, as perceived by the Japanese listener. There is typically much disagreement amongst Japanese lexicographers on the romanji or kana representations of English place names. This is probably because lack of exposure to the spoken form promotes greater reliance on the (highly irregular) English orthographic representation.

With respect to the theoretical goal of the study, an anonymous reviewer made the following astute observation: “If the intent is to ‘investigate the learning mechanisms required for phonological parsing in loanword formation,’ then training on a corpus of loanwords is a surprising choice, since it is usually assumed that loanword phonology is not learned separately but is a side effect of having trained on the internal phonology of the target language”. In other words, perhaps there is an irreconcilable conflict between the engineering and scientific goals of the study. The most direct approach to the problem from an engineering perspective is to train the neural network to perform the mapping between English phonemic and Japanese romanji representations. But this seems clearly the *wrong* approach from the perspective of psycholinguistic modeling of loan word formation, where we postulate an initial stage of segmental phonetic mapping, subsequently constrained by the word prosody of the borrowing language.

It is certainly true that our NN modeling has at least simplified some of the processing which we hypothesise takes place, firstly in accepting as input (American) English phonemic representations, where our model postulates quasi-phonemic segmental representations in the *borrowing* language. We have, in other words, fudged on a level of phonetic to phonemic segmental mapping. This has had a demonstrable, but minor impact on the accuracy of the network performance.

However, the more serious objection that the network should have been trained on native Japanese words may be addressed in the following way. From the perspective of constraint satisfaction, we require an input that highly over-generates with respect to the target forms, but which nevertheless contains all the segmental features to which the output should strive to be *faithful*. We have not attempted to directly simulate the generative capacity of GEN, but rather, provided a mechanism which ensures that segmental faithfulness is met, within the over-generative capacities of the three segment window of the NN architecture. In other words, we argue that by training on the English-Japanese mapping, we have indirectly or approximately simulated the over-generative capacity of GEN, while providing the network with all (and only) the segmental phonetic input to which it is required to be faithful. It is not clear to us how this could be accomplished by training exclusively within the target language.

However, even if this argument carries weight, the second theoretical leg of this study remains to be undertaken: the construction of an analytical (rule-based) competitor to the NN model, and the systematic testing of both against the behaviour of native Japanese speakers' intuitions. We are grateful for the assistance of Chiharu Tsuratani in providing native speaker assessments of the NN responses. This testing needs to be more systematically pursued in further investigations.

Acknowledgments

This work was partially funded by a University of Queensland Postdoctoral fellowship.

References

Bailey, D.C., 1962. *A Glossary of Japanese Neologisms*, University of Arizona Press.

Ingram J. C. (1998) Word Prosody, lexical

access
and the perception of foreign loan words. In J.Wiles (Ed) *Perspectives on Cognitive Science II: Theories, experiments, and foundations,* Ablex: [In press]

Lovins, J. B.(1975) *Loanwords and the phonological structure of Japanese,* Bloomington : Indiana University Linguistics Club.

Rumelhart, D.E., G.E. Hinton & R.J. Williams, 1986. Learning representations
by back-propagating errors, *Nature* **323**, 533-536.

Sejnowski, T.J. & C.R, 1987. Rosenberg, Parallel Networks that Learn to Pronounce English Text, *Complex Systems* **1**, 145-168.