CENTER FOR RESEARCH IN LANGUAGE

January 1998 Vol. 11, No. 3

The Newsletter of the Center for Research in Language,
University of California, San Diego
WWW: http://crl.ucsd.edu/newsletter.html

FEATURE ARTICLE

Lexicons in Contact: A Neural Network Model of Language Induced Change

Lucy Hadden
Department of Cognitive Science, UCSD

EDITOR'S NOTE

This newsletter is produced and distributed by the CENTER FOR RESEARCH IN LANGUAGE, a research center at the University of California, San Diego that unites the efforts of fields such as Cognitive Science, Linguistics, Psychology, Computer Science, Sociology, and Philosophy, all who share an interest in language. We feature papers related to language and cognition distributed via the World Wide Web and welcome response from friends and colleagues at UCSD as well as other institutions. Please visit our web site: "http=://crl.ucsd.edu".

SUBSCRIPTION INFORMATION

If you know of others who would be interested in receiving the newsletter, you may add them to our email subscription list by sending an email to listserv@crl.ucsd.edu with the line "subscribe email-address newsletter" in the body of the message (e.g. subscribe jdoe@ucsd.edu newsletter).
Please forward correspondence to:

Paul Rodriguez, Editor
Center for Research in Language, 0526
9500 Gilman Drive, University of California, San Diego 92093-0526
Telephone: (619) 534-2536 E-mail: editor@crl.ucsd.edu

Back issues of this newsletter are available from CRL in hard copy as well as soft copy form. Papers featured in previous issues include the following:

PDP Learnability and Innate Knowledge of Language
David Kirsh
Department of Cognitive Science, UCSD
Vol. 6, No. 3, December 1991

Why It Might Pay to Assume That Languages Are Infinite
Walter J. Savitch
Department of Computer Science, UCSD
Vol. 6, No. 4, April 1992

Mental Spaces and Constructional Meaning
Claudia Brugman
Center for Research in Language, UCSD
Vol. 7, No. 1, October 1992

Is Incest Best? The Role of Pragmatic Scales and Cultural Models in Abortion Rhetoric
Seana Coulson
Department of Cognitive Science, UCSD
Vol. 7, No. 2, January 1993

Marking Oppositions in Verbal and Nominal Collectives
Suzanne Kemmer
Department of Linguistics, UCSD
Vol. 7, No. 3, September 1993

Abstract Better Than Concrete: Implications for the Psychological and Neural Representation of Concrete Concepts
Sarah D. Breedin, Eleanor M. Saffran, and H. Branch Coslett
Center for Cognitive Neuroscience, Department of Neurology, Temple University
Vol. 8, No. 2, April 1994

Analogic and Metaphoric Mapping in Blended Spaces: Menendez Brothers Virus
Seana Coulson
Cognitive Science, UCSD
Vol. 9, No. 1, February 1995

In Search of the Statistical Brain
Javier Movellan
Department of Cognitive Science, UCSD
Vol. 9, No. 2, March 1995

Connectionist Modeling of the Fast Mapping Phenomenon
Jeanne Milostan
Computer Science and Engineering, UCSD
Vol. 9, No. 3, July 1995

Representing the Structure of a Simple Context-Free Language in a Recurrent Neural Network: A Dynamical Systems Approach
Paul Rodriguez
Department of Cognitive Science, UCSD
Vol. 10, No. 1, October 1995

A Brain Potential Whose Latency Indexes the Length and Frequency of Words
Jonathan W. King
Cognitive Science, UCSD
Marta Kutas
Cognitive Science and Neurosciences, UCSD
Vol. 10, No. 2, November 1995

Bilingual Memory: A Re-Revised Version of the Hierarchical Model of Bilingual Memory
Roberto R. Heredia
Center for Research in Language, La Jolla, CA
Vol. 10, No. 3, January 1996

Development in a Connectionist Framework: Rethinking the Nature-Nurture Debate
Kim Plunkett
Oxford University
Vol. 10, No. 4, February 1996

Rapid Word Learning by 15-Month-Olds under Tightly Controlled Conditions
Graham Schafer and Kim Plunkett
Experimental Psychology, Oxford University
Vol. 10, No. 5, March 1996

Learning and the Emergence of Coordinated Communication
Michael Oliphant, John Batali
Deptartment of Cognitive Science, UCSD
Vol. 11, No. 1, February, 1997

Contexts That Pack a Punch: Lexical Class Priming of Picture Naming
Kara Federmeier and Elizabeth Bates
Cognitive Science, UCSD
Vol. 11, No. 2, April, 1997

Lexicons in Contact: A Neural Network Model of Contact-Induced Language Change

Lucy E. Hadden [1]
Department of Cognitive Science
University of California, San Diego

Abstract

Languages often interfere with one another through contact, with effects which are often large and sometimes surprising. Many theories have been advanced, but, because of the impossibility of running controlled experiments, it is difficult if not impossible to gauge the relative importance of various factors with any degree of accuracy. Such controlled experimentation is, however, possible with a model. I present a neural network community model which is capable of reproducing "Finglish", or American Finnish, which arose in the United States in the early years of the 20th century. This model is suggested as one which could ultimately be used to explore the relative importance of various factors which have been suggested as influencing the outcomes of language contact situations.

1 Background

In the early 20th century, a moderate number of Finns moved to the United States, and soon afterwards the state of their Finnish was deplored in newspapers in Helsinki. One reason was the prevalence of words borrowed from English in sentences which were otherwise (syntactically, phonetically, morphologically) Finnish. In some cases, the entire sentence might be composed of loan-words, as in "Pussaa[p] peipipoki petiruumasta kitsiin" (Push the baby buggy out of the bedroom into the kitchen) (Karttunen, 1977). In fact, the blending of English and Finnish was so great in these immigrants' speech that their Finnish became known as American Finnish, or "Finglish".

Languages can come into contact for a variety of reasons, from shared borders and trading to migration of groups of people into areas where the natives speak another language. Contact can have two major kinds of effects: either speakers of one language borrow structures, such as words, inflections or inflectional systems, phonemes, or word order, from the other (source) language; or they can abandon their own language and adopt the new (target) one. In some cases, such a shift can disrupt the target language, though the circumstances under which this happens are not clear. Borrowing and shift are not mutually exclusive, and often appear in combination. For example, the speakers of Finglish borrowed English words into their Finnish, in the overall context of a shift to English.

Theorists disagree about what determines the outcome of a language contact situation. It may be that contact only acts as a catalyst, speeding up language-internal historical processes (Karttunen, 1977). Or the outcome may depend entirely on the formal structure of the two languages (Weinreich, 1953). According to Weinreich (Weinreich, 1953), the outcome depends on a large number of social and psychological factors, including school language and individual speakers' (particularly bilinguals') attitudes towards the two languages. A fourth approach says that the outcome of language contact situations depends largely on certain aspects of the social structure, and, subsidiary to that, on the structures of the languages involved. For example, (Thomason, 1988) theorize that the intensity of contact between two languages is the major determinant of the result, especially in borrowing situations. They define "intensity" as a combination of the temporal duration of contact, relative numbers of source-language speakers and borrowing-language speakers, and the socio-political dominance of source-language speakers (prestige effects, essentially). The longer the contact continues, the more borrowing will occur, they claim. Similarly, having many more source- language speakers will tend to encourage lots of borrowing, while having many more borrowing-language speakers will tend to restrict the amount of borrowing which occurs. Furthermore, the dominance of the source- language speakers will tend to encourage borrowing, all things being equal, while dominance of the other group will tend to inhibit borrowing.

Unfortunately, although we know about or have evidence of many language contact situations, we often do not have historical information about the socio-cultural relations of the two groups at the time of contact. Furthermore, even in the cases where we do have such information, the situations tend to vary along several dimensions at a time, rather than along a single one. Since historical linguists are restricted to studying past contacts (and even socio-linguists can only work with sit-

[1] Supported by a NPSC Fellowship

uations in progress), controlled experiments are clearly impossible. As a result, the information we have about language contact is not particularly conducive to determining the actual tradeoffs between factors, or even to determining what factors really are important and which are not.

Fortunately, an appropriate model is conducive to exactly those tasks which are hardest within historical and socio-linguistics. A model can easily be manipulated along a single dimension at a time, making controlled comparisons possible. With a model, it may be possible to gain a better understanding both of the tradeoffs between the various factors and of the relative importance of the factors. Before a model can be used for such explorations, however, it must be validated on real language contact situations. That is, a model which can productively be used to understand the factors involved in language contact outcomes must be able to produce results qualitatively similar to real language contact outcomes.

1.1 Overview

This paper proposes such a model, and explores its ability to replicate two historical contact situations: contact between Old French and Old English after the Norman Conquest, and between English and Finnish in the 20th-century United States. The model itself is based on the ones described in Hare & Elman (1992) and (Hare & Elman, 1993), though it uses communities of artificial neural networks to represent the two populations, rather than single exemplars. It begins to examine the relationships between the factors that Thomason (1988) postulate as relevant to language contact outcomes. The chosen examples differ primarily along one dimension which Thomason (1988) deem important: socio-political dominance; this is discussed further below (see sect. 1.3).

1.2 Previous Models

One way to model language-related phenomena is to use neural networks. Hare & Elman (1992) showed that using the output of one network as the teaching signal for another can replicate historical linguistic phenomena. In their work, the networks performed a morphological task, in which the inputs consisted of two banks: one representing the basic verb (essentially semantics, or some such), and the other representing the inflection that was to be given to the verb. The output was the inflected form of the stem, including a kind of "phonology", so that different classes of "stems" could be trained to take different inflectional endings. ?) described the behavior of populations of auto-associative networks. These models were the basis for the current model.

1.3 The Examples

Contact between English and Finnish and between Old English and Old French differed chiefly along one dimension: dominance. As a result, this pair of language contact examples are of particular interest for modeling. Both involved small groups of immigrants, but in the English-Finnish case the small group was subordinate, while in the Old English-Old French case, the small group was decidedly dominant. Their outcomes also differed: Finnish was rapidly abandoned by the Finnish speakers, and in the meantime their Finnish borrowed large numbers of lexical items from English, the dominant language (Karttunen, 1977). In contrast, Old English borrowed a substantial number of words from Old French, and Old French remained in use for several generations (Thomason, 1988).

2 Method

2.1 The mappings

In order to simulate language contact using neural networks, simulated languages are necessary. For these simulations, the "languages" consisted of two constructed mappings, in the style of a "morphological task". The mappings took a set of "concepts", or semantic information (e.g. "pet which barks"), plus information about number (e.g. "plural"), case (e.g. "nominative", or "subject"), and language (e.g. "English"), and produced an inflected word (e.g. "dogs"). (Or, with the same information but language "Latin", produced the word "canes".)

The inputs were 34 bits long, with 16 bits representing the "concept", 4 bits representing number, 4 representing case, and 10 representing the language of the output. (It made sense to give the language information 10 bits, and so 10 weights, to make the original mappings easier to learn, and allow for more emphasis on the identity of the language to be produced during contact. Redundant inputs are frequently helpful in training neural networks.) Each 16-bit "concept" was randomly generated, with a .25 chance that each bit would be on, so that each pattern had an average of 4 bits on. There was no attempt to cluster the concepts. The outputs consisted of 4-phoneme stems, each of which corresponded to one input "concept" for that language, and a 2-phoneme suffix which encoded number and case information. Each phoneme consisted of 7 bits, which encoded various features of those phonemes. Although the phonemes were coded with respect to actual phonetic features (consonant/vowel, place of articulation, etc.), and attempts were made to make the phoneme sets relevant to the languages being modeled, the practical result was abstract "phonemes" and "words" rather than actual subsections of the languages in question.

Each language consisted of 13 concepts and their associated stems, each of which took four inflections (all possible combinations of two numbers and two cases). The mappings differed completely on the output stems and inflections, but shared 9 of the 13 input "concepts" and all of the input representations of number and case information. Thus, out of a total of (13 concepts x 4 inflections =) 52 inputs in each language, 36 were common to both, and each had an additional 16 of its own. Each language had only one declension, or set of suffixes, so that the mapping was fairly simple.

2.2 Initialization

Overall, the simulation consisted of 20 networks, divided into two groups of ten. Each group was assigned one of the two mappings as its "native language", and each network in the group was trained on that mapping before contact began.

The networks were standard three-layer backpropagation networks, with no recurrent connections. The networks had 34 input units, 20 hidden units, and 42 output units, all fully interconnected. There was no bias unit, nor were there any direct connections from the input units to the output units. The outputs for both languages appear in the same set of units. This architecture essentially assumes that bilingualism uses the same language structures and abilities for both languages. (Another architecture, with two sets of outputs, exemplifying the view that bilinguals use different sets of resources for each language, was constructed but will not be discussed in this paper, because the results were less interesting and space does not allow it.)

Each network was trained until the sum squared error over its output nodes was less than 0.5 over 100 inputs presented in random order; in practice, this meant that almost all of the outputs were on the correct side of 0.5, and many were much closer to target than that. Usually, training took between 2000 and 6000 iterations, where each iteration was a presentation of a single, randomly picked input. All networks were trained using a learning rate of 0.3 and momentum of 0.2. The initial random weights were constrained to be less than 0.5.

2.3 Contact

Once all the networks had been initialized, the two groups were brought into contact. Contact consisted of a sequence of interactions between two networks. In each interaction, one of the networks was chosen to act as the "hearer". Then another network was chosen to act as "speaker", and an input from the speaker's native language selected randomly. The input was first presented to the speaker, which produced output, but was not trained at all based on that output. The same input was presented to the hearer, and the output of the speaker network used as the teaching signal for the hearer.

Presenting the same input to both networks assumes that both speaker and hearer have some idea of what the speaker will refer to next, what the word's role in the sentence is, and what the speaker's native language is. All of these assumptions are somewhat reasonable: word identity, case, and number cues could be provided by the sentence context, language information by the identity of the speaker.

The three factors which Thomason and Kaufman describe as relevant to language contact outcomes: duration of contact, relative sizes of the populations, and socio-political dominance by one group, all had to be implemented in this model. For these simulations duration of contact was simply the length of the simulation. Contact continued for approximately 4000 interactions. Each network acted as the hearer exactly 200 times (the program cycled through the networks in order as hearers), and as the speaker an average of 200 times (the speaker was chosen at random from one population or the other each time). Relative population size was implemented as the probability that the speaker came from each group, and was represented by the size ratios. Thus, if the size ratios were .4 and .6, the speaker was picked from the smaller group with probability .4. After the speaker's population had been chosen at random, a specific network was randomly selected to act as the actual "speaker," avoiding the hearer if they were from the same population.

Socio-political dominance is not obviously relevant to neural networks, so its implementation took some thought. In Thomason (1988) formulation, socio-political dominance tends to mean that the dominant group's language interferes more with the non-dominant group's language than vice versa. A possible analogue in neural networks is the learning rate. Thus for this simulation, socio-political dominance was implemented as different learning rates for the two languages. If the speaker came from the dominant group, the learning rate used on the hearer was higher than that used if the speaker came from the subordinate group, regardless of the origin of the hearer. A typical pair of learning rates was 0.2 or 0.25 for the subordinate group and 0.3 for the dominant group.

2.4 The simulations

Two sets of simulations were run, one for each language contact situation. In the first one, simulating Finnish-English contact, the small group was also the one with the lower "prestige" or "dominance", which is to say, inspiring a lower learning rate than the other group. The other set of simulations was for the Old French-Old English situation. In this case, the smaller group had the higher "dominance" or learning rate. For the most part, one mapping was associated with the smaller group; hence, it will be called "Finnish" or "Old French" throughout. The mapping typically associated with the larger group of networks will be referred to as "English." A few simulations were run in which the larger population used "Old French"/"Finnish," to ensure that the mappings themselves were not influencing the outcome too strongly.

In each simulation, a number of different size ratios and learning rate pairs were used, since the exact parameters had to be set by hand. The learning rates differed by no less than 0.05, and no more than 0.1. In addition to various pairs of learning rates, three different size ratios were also tried: 0.25 and 0.75, 0.1 and 0.9, and 0.4 and 0.6.

2.5 Possible Outcomes

The simulations could produce several kinds of results. Ideally, they would replicate the observed outcomes, which are spelled out in more detail for each simulation set. Another possibility was a "creole" of some sort, in which all the networks converge to some mapping that is different from both the original mappings, but is a coherent mapping in and of itself. A third possibility was, of course, catastrophic interference. In this case, the networks would be unable to perform either mapping, and would wind up with gibberish.

2.5.1 Simulation 1: Finnish-English

Seven separate simulations were run in the "Finnish-English" condition. For all these simulations, the mappings fit their standard roles; that is, "Finnish" was the language of the small, subordinate group and "English" was the language of the large, dominant group.

To replicate the observed results, the "Finnish" population will wind up with mostly "Finnish" inflections, but also with a fairly large number of stems taken from "English" even when the input indicates that the output should be in the native language. The "English" mapping should be more or less unaffected, as English was essentially unaffected by Finnish. A small amount of change to the "English" mapping would be acceptable (see Figure 1, "finnish hypoth." columns). Since the interesting outcome of the contact between Finnish and English showed up within one generation, it is quite plausible that if the model works, the effects should be quite similar to the effects seen in the real world.

In these simulations, catastrophic interference should hit the "Finnish" networks first, since they will be seeing the other mapping much more often than their own, and will be using a greater learning rate for the other mapping than they are for their own.

2.5.2 Simulation 2: Old French-Old English

Four simulations were run with the mappings in the standard configuration (smaller "Old French", larger "English"). The "French-speaking" networks belonged to the small but dominant group, while the "English-speaking" networks belonged to the large but subordinate group. For the most part, the same pairs of learning rates were used as in the Finnish-English simulations, but the low learning rate now went to the "English-speaking" networks instead of those speaking "Old French"/ "Finnish".

In this case, behaving similarly to the real contact situation means that the "English" mapping adopts a reasonable number of "French" words, and, ideally, adopts a "French" inflection or so.

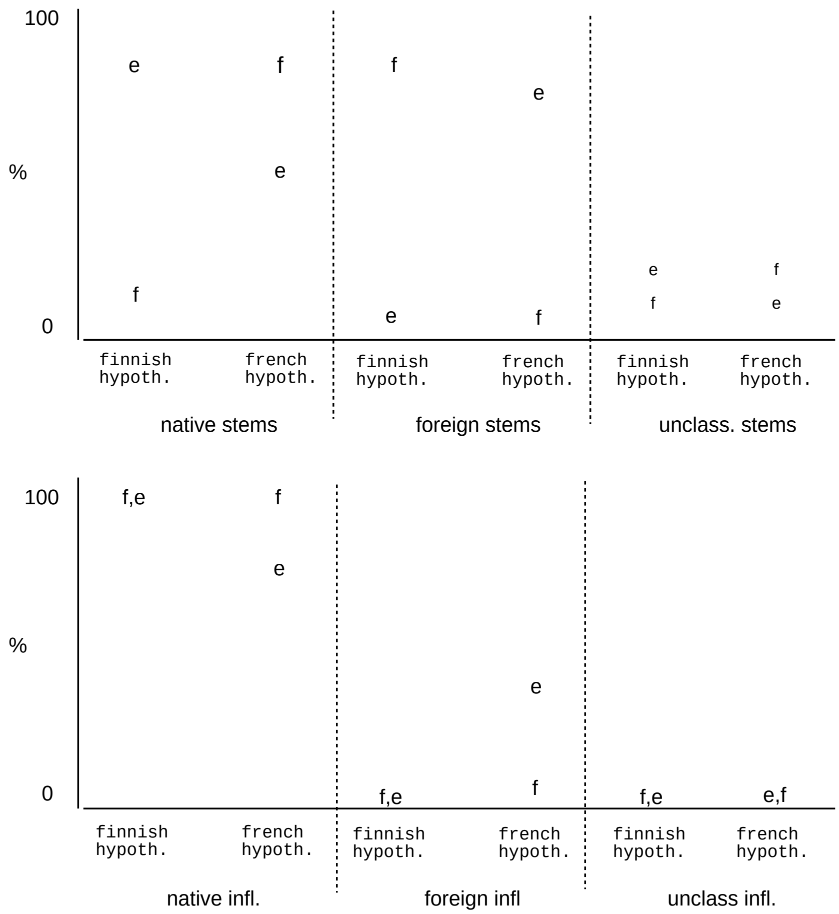

Figure 1: Hypothesized outcomes of simulations. f=finnish (finnish hypoth. columns) or french (french hypoth. columns), e=english. These are roughly the outcomes of the real contact situations. Note that for the networks, unchanged means 76% native stems and 24% unclassifiable stems, because each "language" left out 4 of the "concept" inputs.

"Old French", in contrast, should degenerate significantly, as the Old French spoken in England during the Middle Ages did. It is, however, unlikely that this simulation will yield results similar to that of the real contact, given that the simulation only uses a single generation of networks, and the impact which Old French had on Old English took several generations (see Figure 1, "french hypoth." columns).

2.5.3 Simulation 3: "backwards" conditions

Because the previous simulations were all run with "Finnish"/ "Old French" as the language of the small group and "English" as the language of the large group, a relatively small number of "backwards" simulations were run as controls. Essentially, these assumed the opposite of what happened in real life: that a small number English-speakers moved to Finland or France, under the analogous dominance conditions (subordinate in Finland, dominant in France). Overall, 5 such simulations were run. Two were in the "Finnish backwards" condition and three in the "French backwards" condition.

If the mappings were in fact neutral, then the results in these conditions would be identical to the results in the corresponding normal or "forward" conditions, but with the labels reversed. If, however, inherent properties of the two mappings influenced the outcomes of the simulations then the outcomes of these simulations should be quite different from those of the corresponding "forward" simulations.

2.6 Evaluation

Once the simulation had finished, each stem and inflection produced by the networks was assigned to one mapping or the other. In talking about outcomes of real language contact, historical linguists usually discuss features like how many words come from the native language and how many from the source language, as well as which ones (more common or everyday words are less likely to get borrowed than are uncommon or technical ones or words for new concepts). Similarly, the phonetic and morphological structures of the language before and after contact are compared. For the networks, such evaluation consisted of determining how many stems and inflections produced by the populations after contact came each of the original mappings. Evaluation was thus a pattern classification problem (Duda, 1973).

Stems and inflections were evaluated separately, to gauge the different effects on "lexicon" and "morphology". The networks' responses to native and non-native concepts were also compared; the new mapping should be easier for the networks to learn on new inputs than on old ones. This comparison was only relevant for the stems, since contact only introduced novel concepts to the input set, and never novel inflections.

Since each output is an array of bits, it can be considered as a vector in high-dimensional space. In particular, each 28-bit stem is a 28-dimensional vector, and each 14-bit suffix is a 14-dimensional vector. The stems of each language should tend to cluster together at least somewhat, and be somewhat separated from the stems of the other (see Fig. 2); the inflections should behave similarly. Thus, classifying the output of a network simply consists of finding the shortest distance between it and the points in each of the original mappings, and assigning it to the mapping to which it is closest (see Fig. 3). Of course, the nearest point must be reasonably close, to distinguish gibberish from outputs which actually belong to one mapping or the other. Thus, the distances to the nearest point in each mapping is compared to a threshold distance; if it is greater than the threshold, the output is classified as belonging to neither language (see Fig. 4). The thresholds were set to 1.0 for inflections and 2.0 for stems. Since stems are twice as long as inflections, this allows for essentially the same amount of disparity per bit in the vectors being compared.

The major evaluator was essentially a nearest-neighbor classifier (Duda, 1973). The distance between the average output of the population and each point in each of the original mappings was found, and the smallest one used as the distance from the point to that mapping. Thus, if the average output for a stem or inflection is av, and the nth point of "Finnish" is F_1, with Euclidean distance between points represented by $||x - y||$, the distance to "Finnish", d_1, is $d_1 = min(||av - F_1||)$, and the distance to "English" is $d_2 = min(||av - E_1||)$.

Then av belongs to "Finnish" if $min(d_1, d_2) = d_1$, and to "English" if $min(d_1, d_2) = d_2$. This does a fairly good job of indicating how close the point lay to each mapping's cluster, since any given element of the mapping may have wandered quite far from its original counterpart, but still be more in the space "belonging" to one mapping than in that "belonging" to the other.

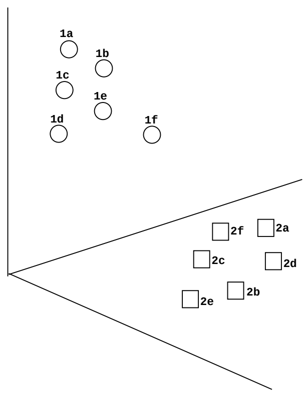

Figure 2: Hypothetical clustering of the stems (or inflections) for the two mappings or "languages."

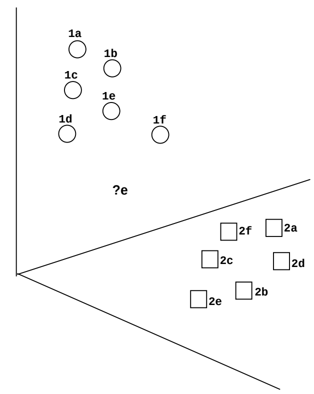

Figure 3: Classifying a network's output as a belonging to a mapping.

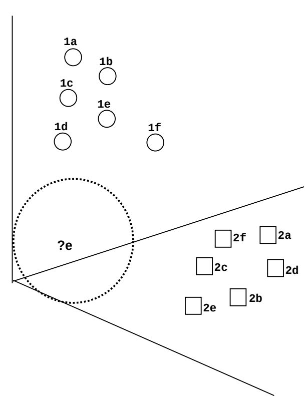

Figure 4: Categorizing outputs that are too far from any cluster as unclassifyable.

3 Results

3.1 Pre-contact results

In order for contact results to be meaningful, it was necessary to test the networks and the classifier before any contact took place. The results were as expected, so that all 13 stems from the network's trained mapping was classified as belonging to that mapping, and the other 4 were classified as belonging to neither (or, occasionally, one was classified as belonging to the other mapping). All 4 inflections were classified as belonging to that mapping. Analysis of the errors on stem and inflection output units also showed that, as expected, the inflections were over-learned when compared to the stems. (Because each inflection appeared 13 times in the training set, and each stem only appeared 4 times, the networks had much more opportunity to learn the inflections than the stems.)

3.2 Finnish-English results

All of the simulations produced outcomes which are qualitatively like the outcome of the actual contact between Finnish and English, as can be seen from the similarity of the "sim." and "hypoth." columns in Figure 5. Essentially, the "English-speaking" networks were unaffected by contact, while the "Finnish" networks produced many "English" stems but mostly used "Finnish" inflections. This is exactly the kind of outcome which was predicted if the contact in fact came out as the real contact between Finnish and English did.

In particular, "English" ended, as it began, with 76% of its stems "native" and the remainder "unclassifiable," and with all its stems "native." "English" was thus unaffected by its contact with "Finnish" just as English was unaffected by the small population of Finnish speakers. "Finnish" also started out with 76% of its stems "native," and ended with 29% "native." The majority, 58%, became "foreign" and the rest (13%) were "unclassifiable." Thus, by the end most of the stems were foreign, just as many words in Finglish were of English origin. The "Finnish" inflections also stayed mostly "native," with 65% "native." The rest were a mix of "foreign" (25%) and "unclassifiable" (10%). So the inflections were much less affected by contact than were the stems, which is precisely the pattern found in Finglish.

3.3 Old French-Old English results

The results of the Old French-Old English simulations never approximated those of the real contact situation. (See Figure 6, "sim." and "hypoth." columns.) Instead, "English" tended to end up with unclassifiable stems and inflections, and no "foreign" ones; "French" did pick up some "foreign" stems and inflections. In at least two cases, an intermediate outcome resembled that of Finnish and English most strongly. In one, "French" picked up a large number of "English" stems but preserved half of its own inflections. "English", on the other hand, deteriorated much more quickly than it had in the Finnish-English simulations, but did not gain any "French" stems.

3.4 Results of the "backwards" conditions

The relatively small number of "backwards" simulations (with "English" as the language of the smaller, less prestigious group) clearly yielded different results from the corresponding normal or "forward" conditions. One of the most striking differences was in the two Finnish-English simulations. In one simulation, both mappings were disrupted roughly equally, and very little. (See Figure 5, "back" column, f1 and e1.) There was almost no learning of non-native stems, and moderate to severe disruptions of the inflections for both mappings. In the other simulation of this type, where the relative sizes of the two populations were almost equal, more disruptions occurred. (See Figure 5, f2 and e2.) For both mappings, and for both stems and inflections, disruption took the form of becoming unclassifiable rather than "foreign."

The "French backward" case produced relatively good results in at least one simulation. Essentially, this took the form of an excellent "Finglish" outcome with much greater simultaneous disruption of "English" than in the actual "Finglish" results. (See Figure 6, "back" columns.)

For evaluation, each network was given all 52 inputs on which it was initially trained, plus the other 16 inputs which it had (presumably) seen during the course of contact. It was expected that the networks should learn these new inputs in the new language more easily than it re-learned old stems, but in fact this was not the case. In the cases where populations adopted foreign words into their lexicons, at most two of those words came from the inputs they had never seen before;

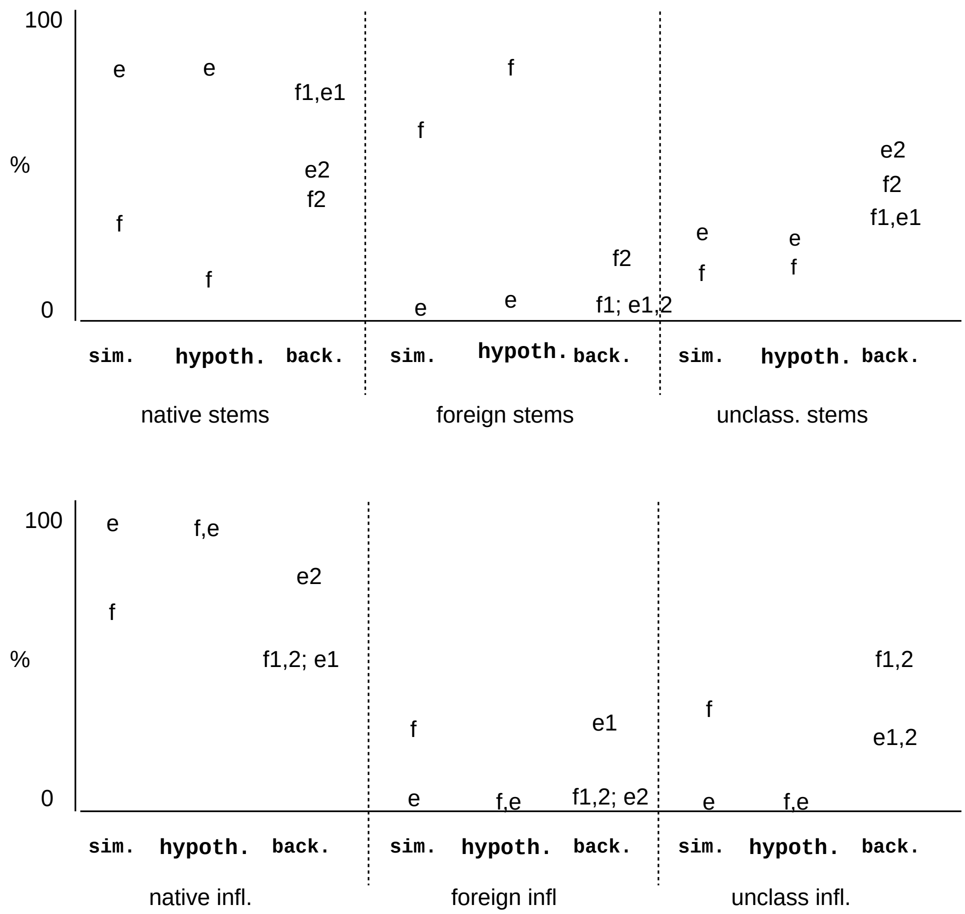

Figure 5: Results of the Finnish-English simulations. f=finnish, e=english. "Hypoth." columns are the desired outcomes (and roughly reflect the outcome of the real contact situations); "sim" columns indicate the average values for the simulations; "back" columns are the outcomes of examples of simulations with the roles of the mappings reversed. (Thus the desired outcome for the "back" columns is for "e" to be in the same place as "f" in the "hypoth." columns, and vice versa.) In all cases, native + foreign + unclass = 100.

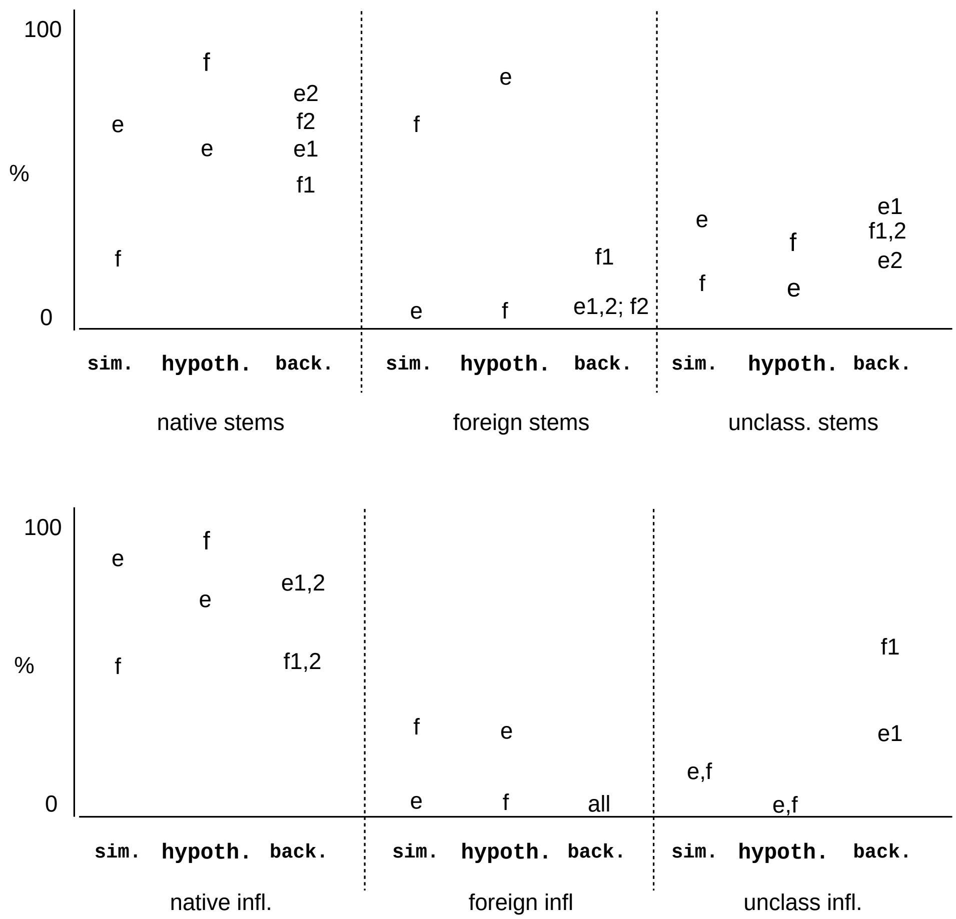

Figure 6: Results of the French-English simulations. f=french, e=english. "Hypoth." columns are the desired outcomes (and roughly reflect the outcome of the real contact situations); "sim" columns indicate the average values for the simulations; "back" columns are the outcomes of examples of simulations with the roles of the mappings reversed. (Thus the desired outcome for the "back" columns is for "e" to be in the same place as "f" in the "hypoth." columns, and vice versa.) In all cases, native + foreign + unclass = 100.

the rest were new stems associated with input patterns which belonged to the networks' native language.

3.5 Summary

The results of the Finnish-English simulations turned out much better than those of the Old French-Old English simulations. All seven Finnish- English simulations produced "Finglish", while only one or two Old French-Old English simulations produced anything like the real Old French-Old English outcome. It was also clear from these simulations that the mappings were not equally susceptible to disruption by each other. Even though they both degraded at about the same rate when allowed to run without any interference from the other mapping, "English" was much more resistant to interference from "Finnish/Old French" than vice versa.

4 Discussion

Although the convergence on "Finglish" in all cases in the "forward" Finnish-English simulations is good, the main finding seems to be that the mappings themselves have at least as strong an effect on the outcomes of the simulations as the factors being tested–differential learning rates (dominance) and differential probabilities of having mappings be taught (size of population). "English" was simply not as susceptible to change as the other mapping.

4.1 "Forward" simulations

Essentially, the network populations in the "Finnish/English" simulations created their own version of Finglish. Many stems were adopted from "English" (analogously with "petiruum" in the real Finglish sentence), but most of the inflections remained "Finnish", as they did in Finglish. Based on this evidence, this model captures some of the features of human language contact well enough to be capable of producing qualitatively similar results.

In contrast, the results of the Old French-Old English contact simulation are quite unlike those found in the real contact situation. "English" never gained any stems from "French", although in fact English owes a large number of its words to its contact with Old French between 9 and 7 centuries ago. This may, however, reflect an oversimplification of the situation rather than an inadequacy of the model. Unlike Finnish in the United States, which vanished within a generation, Old French and Old English remained in contact for generations, and English had an effect on the French spoken in England. Since the results in the real situation took multiple generations to surface, their failure to appear in a single generation of neural net simulations may not be a terrible failure of the model.

In addition to the generational inadequacy of the model, this simulation utterly fails to take into account the possibility of outside teachers for the French speakers in this situation. While the effects of Old French on Old English began to be shown most strongly only after the French speakers had lost their holdings in France and become permanent residents of England (Thomason, 1988, p. 268), French remained a very important language of literature and diplomacy. As a result, the French speakers in England almost certainly heard and read at least a bit of perfectly good Old French from the mainland. In this simulation, however, there is no comparable source of outside information, and so it is, perhaps, not surprising that the results of the simulation do not match the observed results very well.

Thus, some of the disparity in the results between the "Finnish" simulations and "Old French" simulations may be due to the differences between the two situations that did not involve dominance, and which were therefore ignored in the model. Clearly, contact situations which differ only along one dimension are difficult if not impossible to find, which makes constructing and validating models that much harder.

4.2 Effects of mappings

As is clear from the results of the "backwards" simulations, the structure of the two mappings makes a large difference in the outcome of the simulations. The first mapping is apparently more susceptible to interference from the second than vice versa, regardless of the relative frequency of each network's exposure to the two. Thus, the results of the simulations, though clearly influenced by the parameters which were under most investigation, relative size and dominance, seem to have been most affected by the structure of the mappings. It is not clear what features of the mappings make them so different, since they were similar or identical in many respects, including word shape and number of phonemes. Presumably some hidden structure of the mappings is responsible.

4.3 Conclusion

The success of the "Finglish" simulations shows that models involving groups of neural networks can successfully account for at least some forms of contact-induced language change. The effects of the mappings on the outcomes of the simulations points simultaneously to the need for care in constructing such mappings, and the possibility that structural considerations may also be important in real linguistic change, despite Thomason and Kaufman's (1988) claims to the contrary. Better mappings should allow further validation of the method, and further exploration of the interactions of the factors which Thomason (1988) describe as primary.

Hare & Elman (1992); Hare & Elman (1993) have shown that the same mechanisms in neural networks which account for a number of developmental linguistic phenomena can also account for historical language change. These simulations suggest that, with some extensions to populations of networks, they may also be able to account for, and help explore, the factors involved in change due to language contact.

5 References

References

Duda, R. O.; Hart, P. E. (1973). *Pattern Classification and Scene Analysis*. New York: John Wiley & Sons.

Hare, M. & Elman, J. (1992). A connectionist account of English inflectional morphology: Evidence from language change. In *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society*.

Hare, M. & Elman, J. (1993). From weared to wore: A connectionist account of language change. In *Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society*.

Karttunen, F. (1977). Finnish in America: A case study in monogenerational language change. In M. Blount, B. G.; Sanches (Ed.), *Sociocultural Dimensions of Language Change*. New York: Academic Press.

Thomason, S. G.; Kaufman, T. (1988). *Language Contact, Creolization, and Genetic Linguistics*. Berkeley: University of California Press.

Weinreich, U. (1953). *Languages in Contact : Findings and Problems*. The Hague: Mouton.