

THE IMPLICATIONS OF BILINGUALISM AND MULTILINGUALISM FOR POTENTIAL EVOLVED LANGUAGE MECHANISMS

DANIEL A. STERNBERG

*Department of Psychology, Cornell University
Ithaca, New York*

MORTEN H. CHRISTIANSEN

*Department of Psychology, Cornell University
Ithaca, New York*

Simultaneous acquisition of multiple languages to a native level of fluency is common in many areas of the world. This ability must be represented in any cognitive mechanisms used for language. Potential explanations of the evolution of language must also account for the bilingual case. Surprisingly, this fact has not been widely considered in the literature on language origins and evolution. We consider any array of potential accounts for this phenomenon, including arguments by selectionists on the basis for language variation. We find scant evidence for specific selection of the multilingual ability prior to language origins. Thus it seems more parsimonious that bilingualism "came for free" along with whatever mechanisms did evolve. Sequential learning mechanisms may be able to accomplish multilingual acquisition without specific adaptations. In support of this perspective, we present a simple recurrent network model that is capable of learning two idealized grammars simultaneously. These results are compared with recent studies of bilingual processing using eyetracking and fMRI showing vast overlap in the areas in the brain used in processing two different languages.

1. Introduction

In many parts of the world, fluency in multiple languages is the norm. India has twenty-two official languages, and only 18% of the population is a native Hindi speaker. Half of the population of sub-Saharan Africa is bilingual as well. Though bilingualism (or multilingualism, as is often the case) has been investigated in some detail within linguistics and psycholinguistics, it has to date received scant attention from researchers studying language evolution. An extremely important issue remains undiscussed. Whatever theoretical framework one chooses to subscribe to, it is clear that the mental mechanisms used for language processing allow for the native acquisition of multiple distinct languages nearly simultaneously. What is not immediately evident is why they can be used in this way.

On the simplest level, there are two opposing possibilities: either the ability to acquire, comprehend and produce speech in multiple languages was selected for or it came for free as a by-product of whatever mechanisms we use for language. In this paper, we consider a number of the contending theories of language evolution in terms of their compatibility with bilingual acquisition. We test one particular type of general learning mechanism, namely sequential learning, which has been considered a potential mechanism for much of language processing. We propose a simple recurrent network model of bilingual processing trained on two artificial grammars with substantially different syntax, and find a great deal of fine-scale separation by language and grammatical role between words in each lexicon. These results are substantiated by recent findings in neuroimaging and eye-tracking studies of fluent bilingual subjects. We conclude that the bilingual case provides support for the sequential learning paradigm of language evolution, which posits that the existence of linguistic universals may stem primarily from the processing constraints of pre-existing cognitive mechanisms parasitized by language.

2. Potential selectionist theories

Research on bilingualism and natural selection is rather scant, thus selectionist theories on the existence of language diversity may be a good starting point for considering how a selectionist might account for the bilingual case. Interestingly, Pinker & Bloom (1990) argue against a selectionist approach to grammatical diversity, stating that “instead of positing that there are multiple languages, leading to the evolution of a mechanism to learn the differences among them, one might posit that there is a learning mechanism, leading to the development of multiple languages.” This argument rests on the conjecture that the Baldwin effect leaves some room for future learning. Because the previous movement via natural selection toward a more adaptive state increases the likelihood of an individual learning the selected behavior, further distillation of innate knowledge is no longer required after a point (e.g. when the probability nears 100%).

Baker (2003) objects to the claim that the idiosyncrasies of the Baldwin Effect account for the diversity of human languages. He argues that the formidable differences in surface structure between languages should not be glossed over by reference to some minor leftover learning mechanisms. Instead, he suggests that the ability to conceal information from other groups by using a language with which they are unfamiliar could drive the creation of different languages. Like Pinker & Bloom, Baker does not directly argue for a

selectionist model of language differentiation as such, but gives a reason for language differentiation after selection for the linguistic ability has already taken place. What both theories are lacking, however, is an explanation for how this language system can not only accommodate language variation across groups of individuals, but also the instantiation of multiple languages within a single individual.

3. Sequential learning and language evolution

An alternative to the selectionist approach to language evolution can be found in the theory that languages have evolved to fit preexisting learning mechanisms. Sequential learning is one possible contender. There is an obvious connection between sequential learning and language: both involve the extraction and further processing of elements occurring in temporal sequences. Recent neuroimaging and neuropsychological studies point to an overlap in neural mechanisms for processing language and complex sequential structure (e.g., language and musical sequences: Koelsch et al., 2002; Maess, Koelsch, Gunter & Friederici, 2001; Patel, 2003, Patel et al., 1998; sequential learning in the form of artificial language learning: Friederici, Steinhauer & Pfeifer, 2002; Peterson, Forkstam & Ingvar, 2004; break-down of sequential learning in aphasia: Christiansen, Kelly, Shillcock & Greenfield, 2004; Hoen et al., 2003). We have argued elsewhere that this close connection is not coincidental but came about through linguistic adaptation (Christiansen & Chater, in preparation). Specifically, linguistic abilities are assumed to a large extent to have “piggybacked” on sequential learning and processing mechanisms existing prior to the emergence of language. Human sequential learning appears to be more complex (e.g., involving hierarchical learning) than what has been observed in non-human primates (Conway & Christiansen, 2001). As such, sequential learning has evolved to form a crucial component of the cognitive abilities that allowed early humans to negotiate their physical and social world successfully.

4. Sequential learning and bilingualism

Distributional information has been shown to be a potentially crucial cue in language acquisition, particularly in acquiring knowledge of a language’s syntax (Christiansen, Allen, & Seidenberg, 1998; Christiansen & Dale, 2001; Christiansen, Conway, and Curtain, in press). Sequential learning mechanisms can use this statistical cue to find structure within sequential input. The input to a multilingual learner may contain important distributional information that

would also be useful in acquiring and separating different languages. For example, a given word in one language will, on average, co-occur more often with another word in the same language than a word in another language. Thus an individual endowed with a sequential learning mechanism might be able to learn the structure of the two languages. We decided to test this hypothesis using a neural network model that has been demonstrated to acquire distributional information from sequential input (Elman, 1991, 1993).

5. A simple recurrent network model of bilingual acquisition

We used a simple recurrent network (Elman, 1991) to model the acquisition of two grammars. An SRN is essentially a standard feed-forward neural network equipped with an extra layer of so-called “context units”. At a particular time step t an input pattern is propagated through the hidden unit layer to the output layer. At the next time step, $t+1$, the activation of the hidden unit layer at time t is copied back to the context layer and paired with the current input. This means that the current state of the hidden units can influence the processing of subsequent inputs, providing a limited ability to deal with integrated sequences of input presented successively. This type of network is well suited for our simulations because they have previously been successfully applied both to the modeling of non-linguistic sequential learning (e.g., Botvinick & Plaut, 2004; Servan-Schreiber, Cleeremans & McClelland, 1991) and language processing (e.g., Christiansen, 1994; Christiansen & Chater, 1999; Elman, 1990, 1993).

Previous simulations of bilingual processing employing simple recurrent networks have come to somewhat opposing conclusions. French (1998) demonstrated complete separation by language and further separation by part of speech. Scutt & Rickard (1997) found that their model separated each word by part of speech, but languages were intermixed within these groupings. The languages differed in their size (Scutt & Rickard’s contained 45 words compared to French’s 24), however both sets contained only declarative sentences and both used only SVO grammars in their main study. We set out to create a simulation that would more realistically test the ability of this sequential learning model to acquire multiple languages simultaneously. To accomplish this, we used more realistic grammars with larger lexicons and multiple sentence types. We also chose grammars that differed in their word order system.

5.1. Languages

We used two grammars based on English and Japanese, which were modeled on child-directed speech corpora (Christiansen & Dale, 2001). Both grammars contained declarative, imperative and interrogative sentences. The two grammars were chosen because of their different systems of word order (SVO vs. SOV). The English lexicon contained 44 words, while the Japanese was slightly smaller (30 words) due to the language's lack of plural forms.

5.2. Model

Our network contained 74 input units corresponding to each word in the bilingual lexicon, 120 hidden units, 74 output units, and 120 context units¹. The network's goal was to predict the next word in each sentence. It was trained on ~400,000 sentences (200,000 in each language). Following French (1998), languages would change with a 1% probability after any given sentence. The learning rate was set to .01 and momentum to .5.

5.3. Results & Discussion

To test for differences between the internal representations of words in the lexicon, a set of 10,000 test sentences was used to create averaged hidden unit representations for each word. As a baseline comparison, the labels for the same 74 vectors were randomly reordered so that they corresponded to a different word (e.g. the vector for the noun X in English might instead be associated with the verb Y in Japanese). We then performed a linear discriminant analysis on the hidden unit representations and compared the results in chi-square tests for goodness-of-fit. Classifying by language resulted in 77.0% accuracy compared to 59.5% for the randomized vectors [$\chi^2(1, n=74)=5.26, p<.05$]. We also created a crude grouping by part of speech. Though nouns, verbs and adjectives were easy to group, there were a number of words that served a more functional purpose in the sentence, such as determiners, common interrogative adverbs (e.g. "when", "where", "why"), and certain pronouns (e.g. "that"). We classified this set as "function" words. This part of speech classification resulted in 48.65% correct classification, compared with 35.14% for the randomized vectors, but this result was not significant

¹ One reviewer asked about the significance of the number of hidden units used in the model. Generally speaking, learning through back-propagation is rather robust to different quantities of hidden units. It is unlikely that choosing any number of hidden units slightly below or even quite a bit above the number of inputs units would yield different results other than on the efficiency of training (in this case the amount of training required to reach a proficient state).

$[\chi^2(1,n=74)=2.78, p=.099]$. When words were grouped by language and part of speech combined (thus creating eight categories), accuracy rose to 68.92%, compared with 17.57% for the randomized version $[\chi^2(1,n=74)=39.8, p<.001]$. These discriminant analysis results indicate that the net places itself in different internal states when processing English and Japanese. Importantly, the network is sensitive to the specific constraints on parts of speech within each language as indicated by the last analysis which demonstrates a highly significant difference between the trained and baseline accuracy.

These results seem to support local-scale language separation rather than the emergence of two completely distinct lexicons. Though the ambiguous “function” grouping might have created some noise in the data, grouping by language and part of speech gave a highly significant result, seeming to imply that the network attends to both language and part of speech, rather than primarily focusing on one.

6. General Discussion

The bilingual case, as the most prevalent form of language fluency in the world, must be considered in any explanation for the existence of human language. We have argued that it seems difficult to develop a selectionist account of bilingualism. In contrast, a theory of language origins and evolution via sequential learning may be more parsimonious in this regard because it seems to account for bilingualism without needing any major post-hoc revisions. Our simulation of bilingual acquisition via sequential learning demonstrated language separation at a very local scale (i.e. within part of speech and language), rather than the creation of two completely separate lexicons. Converging evidence from neurological and low-level perceptual studies of bilingual processing seem to support this finding. Recent neuroimaging data points to a great deal of overlap in the brain areas used to process different languages in fluent bilinguals (Chee et. al, 1999a, 1999b; Hasegawa et. al, 2002). Eye-tracking studies of fluent bilinguals have also demonstrated partial activation for phonologically-related words in a language not used in the experimental task (Spivey & Marian, 1999)

There are many aspects of language that need to be considered in a final model of bilingual acquisition that were not included in our first model. However, there are at the moment few contending explanations for how this ability came to exist. Our work thus far serves as a first step in demonstrating that sequential learning might be able to account for the ability to process not

only a single language as shown in previous work, but also the ability to process multiple languages simultaneously.

Acknowledgements

We thank Rick Dale for providing his *sentgen* script as well as his English and Japanese grammars, which were used to create the sentences in the simulation. We also thank Luca Onnis and three anonymous referees for their helpful comments and feedback on earlier drafts of this paper.

References

- Baker, M.C. (2003). Linguistic differences and language design. *Trends in Cognitive Sciences*, 7, 349-353.
- Botvinick, M., & Plaut, D. C. (2004). Doing without schema hierarchies: A recurrent connectionist approach to normal and impaired routine sequential action. *Psychological Review*, 111, 395-429.
- Chee, M.W.L., Tan, E.W.L., & Thiel, T. (1999). Mandarin and English single word processing studied with functional magnetic resonance imaging. *Journal of Neuroscience*, 19, 3050-3056.
- Chee, M.W.L., Caplan, D., Soon, C.S., Sriram, N. Tan, E.W.L., Thiel, T., & Weekes, B. (1999). Processing of visually presented sentences in Mandarin and English studied with fMRI. *Neuron*, 23, 127-137.
- Christiansen, M.H., Allen, J. & Seidenberg, M.S. (1998). Learning to segment speech using multiple cues: A connectionist model. *Language and Cognitive Processes*, 13, 221-268.
- Christiansen, M.H. & Chater, N. (Eds.). (2001). *Connectionist Psycholinguistics*. Westport, CT: Ablex.
- Christiansen, M.H. & Chater, N. (in preparation). *Language as an organism: Language evolution as the adaptation of linguistic structure*. Unpublished manuscript, Cornell University.
- Christiansen, M.H., Conway, C.M. & Curtin, S.L. (in press). Multiple-cue integration in language acquisition: A connectionist model of speech segmentation and rule-like behavior. In J.W. Minett & W.S.-Y. Wang (Eds.), *Language Evolution, Change, and Emergence: Essays in Evolutionary Linguistics*. Hong Kong: City University of Hong Kong Press.
- Christiansen, M.H. & Dale, R. (2001). Integrating distributional, prosodic and phonological information in a connectionist model of language acquisition. In *Proceedings of the 23rd Annual Conference of the Cognitive Science Society* (pp. 220-225). Mahwah, NJ: Lawrence Erlbaum.

- Christiansen, M.H., Kelly, L., Shillcock, R., & Greenfield, K. (2004). *Artificial grammar learning in agrammatism*. Unpublished manuscript, Cornell University.
- Conway, C. M., & Christiansen, M. H. (2001). Sequential learning in non-human primates. *Trends in Cognitive Sciences*, 5(12):539--546.
- Elman, J.L. (1990). Finding structure in time. *Cognitive Science*, 14, 179-211.
- Elman, J.L. (1993). Learning and development in neural networks: The importance of starting small. *Cognition*, 48, 71-99.
- French, R.M. (1998). A simple recurrent network model of bilingual memory. In *Proceedings of the 20th Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Lawrence Erlbaum.
- Friederici, A.D., Steinhauer, K., & Pfeifer, E. (2002). Brain signatures in artificial language processing. *Proceedings of the National Academy of Sciences*, 99, 529-534.
- Hasegawa, M., Carpenter, P.A., & Just, M.A. (2002). An fMRI study of bilingual sentence comprehension and workload. *Neuroimage*, 15, 647-660.
- Hoen, M., Golembiowski, M., Guyot, E., Deprez, V., Caplan, D., & Dominey, P.F. (2003). Training with cognitive sequences improves syntactic comprehension in agrammatic aphasics. *NeuroReport*, 495-499.
- Koelsch, S., Schroger, E., & Gunter, T.C. (2002). Music matters: preattentive musicality of the human brain. *Psychophysiology*, 39, 38-48.
- Maess, B., Koelsch, S., Gunter, T., & Friederici, A.D. (2001). Musical syntax is processed in Broca's area: an MEG study. *Nature Neuroscience*, 4, 540-545.
- Marian, V., Spivey, M.J., & Hirsch, J. (2003). Shared and separate systems in bilingual language processing: Converging evidence from eyetracking and brain imaging. *Brain and Language*, 86, 70-82.
- Patel, A.D. (2003). Language, music, syntax and the brain. *Nature Neuroscience*, 6, 674-681.
- Patel, A.D., Gibson, E., Ratner, J., Besson, M., & Holcomb, P.J. (1998). Processing syntactic relations in language and music: an event-related potential study. *Journal of Cognitive Neuroscience*, 10, 717-733.
- Petersson, K.M., Forkstam, C., & Ingvar, M. (2004). Artificial syntactic violations activate Broca's region. *Cognitive Science*, 28, 383-407.
- Pinker, S., & Bloom, P. (1990). Natural language and natural selection. *Behavioral and Brain Sciences*, 13, 707-784.
- Scutt, T., & Rickard, O. (1997). Hasta la vista, baby: 'bilingual' and 'second-language' learning in a recurrent neural network trained on English and Spanish sentences. In *Proceedings of the GALA '97 Conference on Language Acquisition*.
- Spivey, M.J. & Marian, V. (1999). Crosstalk between native and second languages: Partial activation of an irrelevant lexicon. *Psychological Science*, 10, 281-284.