Chapter 7
Monolingual Biases in Simulations of Cultural Transmission

Seán Roberts

7.1 Introduction

Cultural groups are very rarely isolated. They interact for trade, politics and war. Communication is key to these interactions, and so a common language is important. The emergence of common languages has been studied using computational models. However, one aspect of cultural interaction has been left largely ignored – the ability to learn many languages at once, or bilingualism. This chapter considers the importance of incorporating bilingualism into studies of cultural evolution.

Bilingualism is by no means a rare phenomenon. Statistics on the exact prevalence of bilingualism are difficult to obtain. In the USA 18 % of the population are estimated to speak two or more languages (U.S. Census Bureau 2003). The estimate is 34 % for Canada (Statistics Canada 2007), 66 % in the EU (Euopean Comission 2006) and 80 % in China (Baker and Jones 1998). Bilinguals are a majority in about a third of countries (Baker and Jones 1998). These are likely to be conservative estimates, and with over 6,000 languages squeezed into in around 200 nations, it's likely that contact with multiple languages is an everyday feature of most people's lives.

Recently, industrialisation and globalisation have meant that, in the first world, the perception of the prevalence of bilingualism is artificially low – especially for native speakers of global languages such as English (Thomas and Wareing 1999; Kostoulas-Makrakis 2001; Luchtenberg 2002). It's no surprise, then, that when cultural processes come to be modeled, one of the first simplifying assumptions would be that people speak one language. However, the abstraction to monolingualism ignores several linguistic phenomena such as the prevalence of bilingualism in societies and the ease with which children learn more than one language (Pearson et al. 1993).

S. Roberts (✉)
Language and Cognition Group, Max Planck Institute for Psycholinguistics,
Wundtlaan 1, 6525 XD Nijmegen, The Netherlands
e-mail: sean.roberts@mpi.nl

This chapter will consider the validity of monolingual assumptions in models of cultural evolution. Firstly, the way in which bilingualism might affect the evolutionary dynamics of language is explored. Next, a case-study of the 'Minimal Naming Game' will reveal an implicit monolingual bias, namely mutual exclusivity (the assumption that each object only has one name and each name only refers to one object, see Markman and Wachtel 1988 and Merriman and Bowman 1989). Since bilinguals do not exhibit mutual exclusivity (Byers-Heinlein and Werker 2009; Healey and Skarabela 2009; Houston-Price et al. 2010), the model is generalised to weaken this constraint. The model demonstrates that communicative success can be achieved even without mutual exclusivity, in opposition to previous research (Smith 2002, 2009). The model suggests that cultural phenomena adapt to the function they are required to fulfill (e.g. Christiansen and Chater 2008 and Beckner et al. 2009). When seeking to model the integration of cultures a common measurement is required. However, even small differences in the way different communities interact can lead to fundamental cultural differences between them, meaning that a common metric might be very abstract.

7.1.1 Bilingualism and Cultural Evolution

The dynamics of language evolution have been extensively studied through computational modelling. The canonical language learner in these models is an agent that tries to settle on a single grammar that explains the variation in its input. This implicit monolingualism is seen as a necessary abstraction in order to get at the more fundamental dynamics of language evolution. There is a sense in the field of language evolution that bilingualism is a sociolinguistic phenomenon that is the product of the interactions of several monolingual communities who have already evolved language. Implicitly, bilingualism is seen as a secondary linguistic ability – a sort of by-product.

For instance, many models represent languages as discrete entities which compete with one another (Niyogi and Berwick 1995; Abrams and Strogatz 2003). Even when language is modelled as distributions over words, two standard simplifying assumptions are made by many approaches to language evolution and change (e.g. Griffiths and Kalish 2007; Kirby 2001 and Smith et al. 2003). Firstly, it is assumed that there are discrete generations with one agent per generation. This limits the amount of complexity that can be added by the cultural system. Secondly, it is assumed that all learners use the same learning algorithm, or that learning algorithms do not change over a learner's lifetime.

The first assumption has already been criticised (Niyogi and Berwick 2009; Burkett et al. 2010) and recent research has shown that the complexity of cultural dynamics can effect the eventual distribution of languages in a population (Smith 2009). A model has also been proposed which allows agents to speak and acquire multiple languages from multiple speakers (Burkett et al. 2010).

However, the second assumption may also be called into question. I will illustrate this with research on the mutual exclusivity bias, and continue in the next section to show that this bias exists in certain models of language evolution and change. It has been demonstrated that monolingual children and adults exhibit a mutual exclusivity bias (Markman and Wachtel 1988; Merriman and Bowman 1989): a tendency to assume that each object only has one name and each name only refers to one object. However, recent research has shown that bilinguals do not exhibit mutual exclusivity (Byers-Heinlein and Werker 2009; Healey and Skarabela 2009; Houston-Price et al. 2010). It is hypothesised that the bias is overridden because of a higher variance in the input of children in bilingual contexts. Applying mutual exclusivity when presented with two languages is not suitable, since there will be at least two words for each object.

If the amount of linguistic variance (at any level of description) influences the learning strategy for that variance, then this will affect the selective pressure on languages. This will, in turn, affect the kinds of languages that emerge, thus feeding back into the amount of linguistic variance. These aspects would then co-evolve.

Given this, there are two possible fundamental states of the language learner. Either they begin with a mutual exclusivity bias which is overridden in certain situations or they begin with no assumptions and develop mutual exclusivity if the conditions are right. In the next section, it will be shown that some models make implicit assumptions about the development of mutual exclusivity and see it more as a fundamental part of language acquisition and language evolution rather than an acquired heuristic that is applied in suitable contexts. It will be argued that the most abstract learner is one without the mutual exclusivity bias, and so models should not assume mutual exclusivity as part of the learner's bias.

7.2 Categorisation Games

This section presents a case-study of a model of cultural evolution – the Categorisation Game – and demonstrates implicit monolingual biases that obscure some interesting dynamics. The Categorisation Game looks at how agents in a population converge on a shared system for referring to continuous stimuli (Nowak and Krakauer 1999; Steels 1996; Steels and Belpaeme 2005). This paradigm is often couched in terms of deciding on words for objects referred to by their colour. The colour spectrum is continuous, so agents must decide where to place category boundaries as well as the label for that category. The 'minimal naming game' (Loreto et al. 2010) (also used in Gong et al. 2008; Puglisi et al. 2008 and Baronchelli et al. 2010) is a simplification of the categorisation game which "possibly represents the simplest example of the complex processes leading progressively to the establishment of complex human-like languages" (Loreto et al. 2010). I'll show that even this 'minimal' algorithm has implicit monolingual assumptions. First, however, a note is made about the measurements that researchers have used to study the categorisation game.
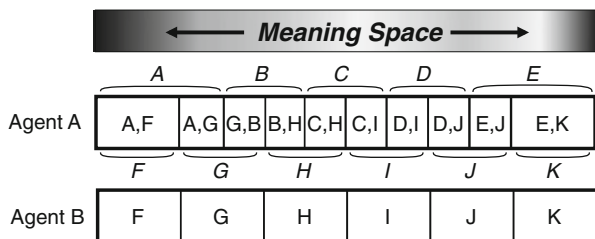
Fig. 7.1 An example of how two agents might split the meaning space into categories and label those categories. The meaning space spans the interval 0.0–1.0. *Agent A* and *B* both have the same conceptual space, but *agent A* has multiple labels in each category while *agent B* only has one label in each category. The representation for *agent A* above pulls apart sections of the space that are contiguously labelled with the same label into two systems

7.2.1 Measurements of Coherence

Other models looking at this problem have considered measurements apart from communicative success. For instance, the 'level of lexical coherence' in the system, according to Baronchelli et al. (2006) is the average proportion of shared lexical items in a population. The category overlap function (Loreto et al. 2010; Puglisi et al. 2008) measures the level of alignment between the category boundaries of the agents. However, an appropriate measurement when considering the possibility of 'bilingualism' is less clear. For instance, consider the example of two agents with categories and labels as described in Fig. 7.1. Adapting the lexical coherence measurement from Baronchelli et al. (2006) gives a coherence of 75 %. This measurement fails to capture the fact that agent B would always be understood by agent A and that agent A could always make itself understood to agent B given the right choice of lexical item. In other words, although the agents have differences in the words that they know, they are still able to communicate unambiguously about the whole spectrum.

Measuring category overlap is also problematic. Agents with category boundaries at exactly the same locations will have a category overlap of 1.0. However, the overlap of the example above is 0.09, despite the relatively good communicative success possible between the pair. This is because the measurement collapses the category boundaries of an agent into a single system before comparing it to another agent. By doing this, the division between the two 'languages' of agent A in Fig. 7.1 is ignored.

These measures reflect the level of coherence in the population, but only effectively for a population whose goal it is to converge on a single, 'monolingual' system. Researchers have used these measurements to gauge the progress of their model, demonstrating a monolingual bias in their approach. Further research is required to find a good way of measuring coherence in a heterogenous population (see Komarova and Jameson 2008; De Vylder 2006 and De Beule 2006). This paper will proceed assuming that communicative success should be the most important measure of coherence between agents.

7.2.2 The Minimal Naming Game

The algorithm for the categorisation game is reproduced below. However, two of the steps are re-analysed as heuristics rather than essential elements. These heuristics impose a mutual exclusivity bias in the agents. The steps are as follows (following Puglisi et al. 2008): There is a population of N agents, each able to partition the perceptual space into categories. Each category has a list of associated words. Each agent has a minimum perceptual difference threshold d_{min}, below which stimuli appear the same. At each time step:

1. Two individuals are chosen at random to be the speaker and the listener.
2. They both have access to a scene containing M stimuli. The stimuli must be perceptually distinguishable by the agents (perceptual distance $\geq d_{min}$).
3. The speaker selects a topic and discriminates it in the following way:

   - Each stimulus is assigned to a perceptual category
   - If one or more other stimuli are assigned to the same category as the topic, the agent splits its perceptual categories so that each stimulus belongs to only one perceptual category. Within a category with two or more stimuli, a boundary is placed halfway between the first two stimuli.
   - The new partitions inherit the associated words of the old partition.
   - **Heuristic A**: Each new partition is given a new, unique name. It's assumed that no two agents will create the same name.

4. The speaker transmits a word that it associates with the topic to the listener. If it has no words associated with the category, it creates a new one. If it has more than one word associated, it transmits the one that was last used in a successful communication.
5. The hearer receives the word and finds all categories which have the associated word and which identify one of the stimuli in the scene. Then:

   - If there are no such categories, the agent does nothing.
   - If there is one such category, the agent points to the associated stimulus.
   - If there is more than one such category, the agent points randomly at an associated stimulus.

6. The hearer discriminates the scene, as above.
7. The speaker reveals the topic to the listener.
8. If the hearer did not point to the topic, the communication is a failure. The hearer adds the transmitted word to the category discriminating the topic.
9. If the hearer pointed to the topic, the communication is a success.
   Heuristic B: Both agents delete all other words but the transmitted one from the inventory of the category discriminating the topic.

Heuristic A, above, invents new words for each sub-category when a category is split. This is an implementation of the assumption that each name only refers to one object, hence when there are two objects with the same name, the agent should

discriminate between them linguistically. This interacts with Heuristic B which removes all competing names associated with a category from the listener's lexicon when communication is successful. The effect is that the listener conforms to the speaker's labeling, but also 'forgets' any previously associated words. This is an implementation of the assumption that each object only has one name.

These two heuristics, then, implement a mutual exclusivity bias: Each name only refers to one object and each object is only labeled by one name. Stable bilingualism is impossible in this model because only one name is retained after successful communication. The role of the two heuristics in the evolution of a shared communicative system is clear: heuristic A creates new labels for categories, introducing variation into the system needed to distinguish between categories. Heuristic B causes the agents to converge on shared labels for categories by selecting for labels common to an interacting pair.

However, these heuristics are still arbitrary. As we have seen, not all human learners assume mutual exclusivity. In the next section, it will be demonstrated that a population of agents can converge on a shared communication system without these heuristics.

7.3   Convergence Without Mutual Exclusivity

The algorithm was modified to remove the mutual exclusivity bias in order to test the effects on communicative success. However, the changes to the dynamics will not be explored in detail. The purpose of the changes, here, is not to explore the best way of modelling the cultural evolution of language, but to demonstrate that the biases of the researcher can influence the dynamics of the model and thus the conclusions drawn from it.

Heuristic B can be modified while retaining communicative success (Baronchelli 2011). If the hearer, but not the speaker applies heuristic B, a coherent vocabulary still emerges in a similar time with similar memory resources required. If only the speaker applies heuristic B a coherent vocabulary does emerge, but on a longer timescale and in a qualitatively different way (approached as a thermodynamic system, consensus is reached due to large, system-size fluctuations of the magnetisation (Baronchelli 2011)). However, this research was concerned with the effect of feedback on the convergence dynamics. This study looks at the assumptions built in to the individual learning algorithm.

The heuristics were modified by generalising the algorithm. Firstly, agents in a population either all applied heuristic A or all did not apply heuristic A. Heuristic B was made optional in the same way. If heuristic B did not apply, a maximum number of words s_{MAX} were retained after a successful communication. A first-in, first-out stack memory was also implemented so that the oldest stored form would be removed first. A word was pushed further back in the stack (safer from deletion) when a listener heard it being used by a speaker. This is a generalisation

of the mechanism that weakens links between signals and meanings which do not co-occur.

The purpose of generalising the model was to allow bilingualism. However, the advantages of knowing more than one word for an object are not yet fully available. A bilingual, failing to communicate with one word, might try another. Therefore, the algorithm was modified to allow an arbitrary maximum number of attempts a_{MAX} at communicating before communication failed. If speakers had more than one label for a perceptual category, they transmitted them in a random order until this maximum was reached. Listeners searched their lexicon at each attempt until either they found a match in their own lexicon and made a guess at the referent or the maximum number of attempts was reached and they signaled failure, as before. Each guess was independent of any other, so successful communication was not always guaranteed, even when $a_{MAX} = M$.

It has been shown that an algorithm which leads to successful communication in a population of agents must strengthen connections between signals and meanings that appear together (or are absent together) and weaken connections between signals and meanings that do not co-occur (Smith 2002). The changes to the algorithm above do not violate these conditions, but simply weaken their strength.

7.4 Results

Four versions of the algorithm were run: with both heuristics, as in the original, with only heuristic A, only heuristic B and with neither heuristic. Results shown here are for a population of 4 over 10,000 rounds with a context size of 2.

7.4.1 Communicative Success

Table 7.1 and Fig. 7.2 show the communicative success for the algorithm run with different heuristics with $a_{MAX} = 2$ and $s_{MAX} = 2$. All heuristics manage in achieving good communicative success at some point (shown by the maximum communicative success achieved). That is, a mutual exclusivity bias is not necessary for communicative success in this model. It should be noted that the probability of choosing the correct referent by chance is $\frac{1}{c} = 0.5$ (where c is the context size) because the algorithm tends to limit the number of words linked to a perceptual category to one. However, algorithms without heuristic B (i.e. 'A only' and 'None') have a higher probability because they are more likely to be able to take advantage of extra communicative attempts. Therefore, for algorithms without heuristic B, the probability of selecting the correct referent by chance is

$$\sum_{i=1}^{a_{MAX}} \left(\frac{1}{c}\right)^i \tag{7.1}$$

Table 7.1 Communicative success for different heuristics for 10 runs each with 4 agents and $a_{MAX} = 2$ and $s_{MAX} = 2$ for 10,000 rounds. The maximum communicative success was calculated as follows: for each run, the average communicative success for data grouped into 100-round bins. The average of a bin was taken as the value of that bin. The values shown in the table are the average of the maximum bin values for 10 runs. The final communicative success is the average success for last 100 rounds over 10 runs. The statistics show a two tailed t-test indicating performance above chance

Heuristics	Max	t	p	Final	t	p
A and B	0.84	23.89[a]	< 0.0001	0.62	3.06[a]	0.01
A only	0.95	44.26[b]	< 0.0001	0.86	7.17[b]	< 0.0001
B only	0.77	9.04[a]	< 0.0001	0.52	1.44[a]	0.18 (N.S.)
None	0.94	43.50[b]	< 0.0001	0.84	5.96[b]	< 0.001

[a]Calculated with chance level at 0.5
[b]Calculated with chance level at 0.75
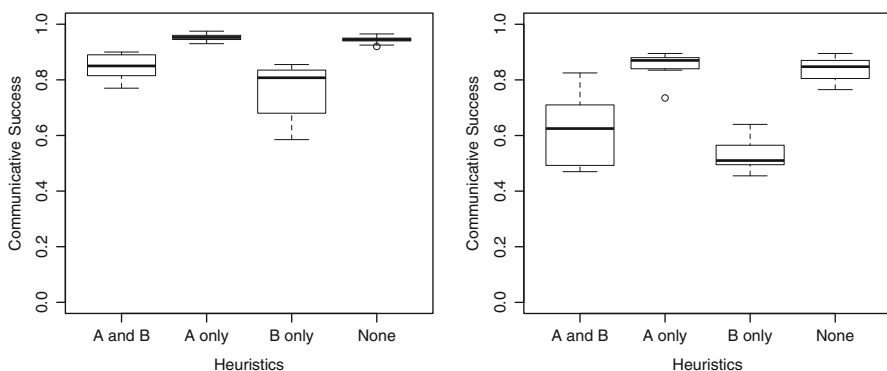
Fig. 7.2 Maximum communicative success (*left*) and final communicative success after 10,000 rounds (*right*) for 10 runs of populations with various heuristics. Statistics in Table 7.1

For the current settings, this is 0.75. Even taking this into account, all algorithms are able to reach stable periods with high levels of communicative success. The result is robust against changes to s_{MAX}: The relative communicative success between the different heuristic combinations remains the same for s_{MAX} up to 1,000, while the absolute communicative success drops about 5 % for s_{MAX} of 4 and remains around that level for s_{MAX} up to 1,000.

However, eventually all agents converge on a single word for the whole meaning space. This is typical behaviour for this model (Baronchelli 2006). This reduces the communicative success, since agents cannot distinguish linguistically between referents. Table 7.1 shows the average final communicative success after 10,000 rounds. These are less than the maximum. In the case of using heuristic B only, the communicative success is no better than chance. The other algorithms still yield a communicative success above chance, but the algorithms without heuristic B (A only and no heuristics) do better than algorithms with heuristic B (average with B = 0.57, without B = 0.85, t = 10.9, p < 0.0001). The same collapsing process
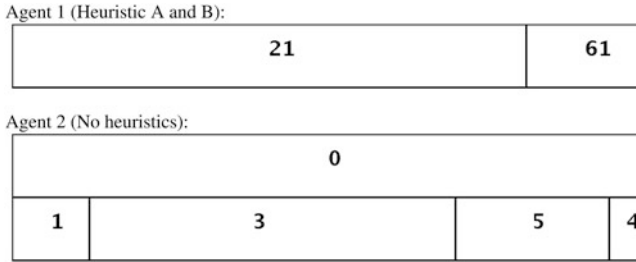
Fig. 7.3 The linguistic labels of an agent after 3,500 rounds for separate runs. Agent 1 (*above*) used heuristics A and B and agent 2 (*below*) used neither heuristic. The perceptual space runs from left to right. Contiguous linguistic categories are indicated by boxes with the linguistic label (a number in this implementation) drawn in the centre. Agent 2 has more than one label for a given perceptual stimulus

occurs as in the algorithm without heuristic B, but since there is more variation within perceptual categories due to extra labels being stored, a single label takes longer to dominate. In fact, a single linguistic item tends to spread over the whole meaning space as with the original algorithm, but a sort of secondary 'language' keeps distinctions between perceptual categories for longer.

Figure 7.3 illustrates this with a diagram of agents' memories from mid-way through separate runs. Agent 1 was run in a population using both heuristics and agent 2 was run in a population using neither heuristic. Agent 1's linguistic categories are already heavily collapsed while Agent 2 has a greater variation which allows it to communicate more effectively. The memories of both agents at this point are nearly perfectly similar to the other agents that they interact with.

Another measure of communicative efficiency is the entropy efficiency of an agent. Effectively, this is the average probability that an agent has a different linguistic label for any two stimuli. An agent has a set of linguistic labels which uniquely identify regions of the meaning space. L is the list of lengths of these regions. The entropy efficiency is given as

$$-\sum_{i=1}^{|L|} \frac{L_i \, log(L_i)}{log(1/d_{min})} \tag{7.2}$$

Since d_{min} is set so that there can be a maximum of 10 perceptually distinct regions, the highest entropy efficiency is given by an agent who can uniquely label 10 regions of equal length (entropy efficiency of 1.0). The lowest possible entropy efficiency is given by an agent with no labels or one label spanning the whole meaning space (entropy efficiency of zero). Figure 7.4 shows that the algorithm with both heuristics achieves a lower entropy efficiency than the algorithm without heuristic B and degrades faster than the algorithm without heuristic A.
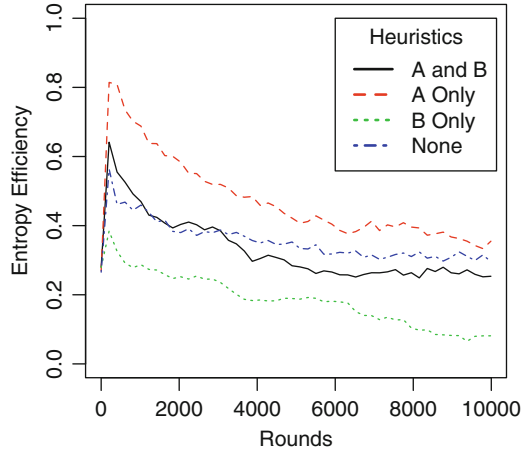
Fig. 7.4 Entropy efficiency for populations of agents with different heuristics. Number of agents = 4, $a_{MAX} = 2$ and $s_{MAX} = 2$
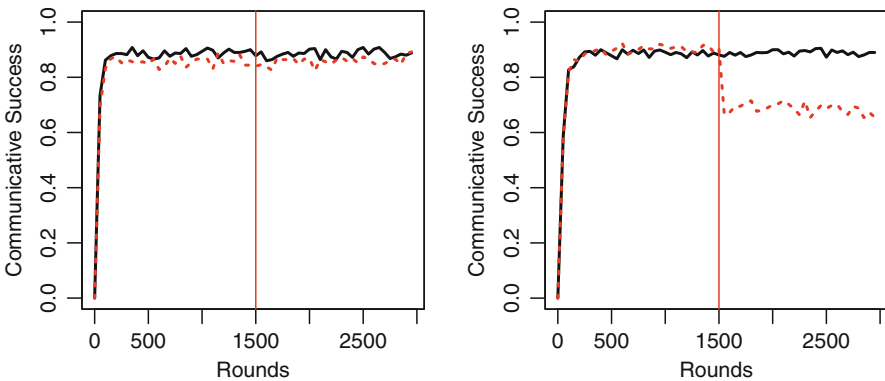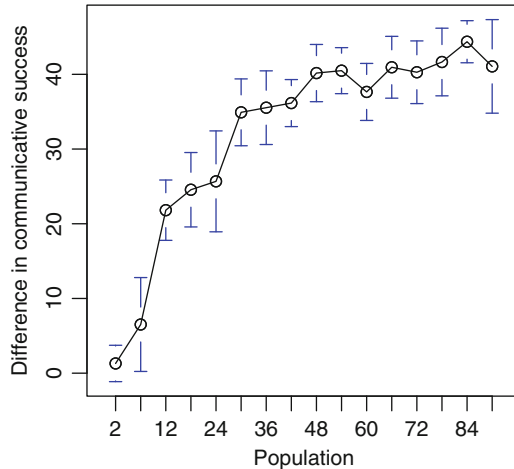

Fig. 7.5 Communicative success for a population of two agents (*left*) and a population of four agents (*right*). *Solid lines* indicate success for a consistent algorithm where no heuristics are applied. *Dashed lines* represent success for an algorithm that incorporates the heuristics after 1,500 rounds

7.4.2 The Development of Mutual Exclusivity

The model has shown that mutual exclusivity is not necessary for communicative success. However, the mutual exclusivity bias is exhibited by monolinguals. The model can be manipulated to explore the rationale behind this and the most likely starting assumptions of a language learner.

Simulations were run where the mutual exclusivity heuristics were 'switched on' after some rounds. Figure 7.5 shows the difference between an algorithm that has no heuristics and one that changes to incorporate them after 1,500 rounds. For

Fig. 7.6 The difference in percentage communicative success when switching from using no heuristics to using both heuristics for different population sizes (the difference between *solid* and *dashed lines* after 1,500 rounds in Fig. 7.5). Values shown are mean and 99 % confidence intervals for five runs. Positive values indicate an advantage for using no heuristics. In a population of 2, there is no difference between using the heuristics or not, as shown in Fig. 7.5

a population of two agents (low cultural complexity), switching on the heuristics makes no difference to the communicative success. Therefore, in this situation, applying mutual exclusivity makes rational sense in order to save memory: The application of heuristic B will reduce the number of words stored for each category. However, in a population of four agents, switching on the heuristics decreases the communicative success. In this situation, the most rational approach is to keep the heuristics switched 'off'. This is because the complexity of the cultural system is greater with four agents, leading to more variation between agents. The system evolves to store many words for an object to cope with this variation. The drop in this difference reflects the empirical findings that bilinguals do not exhibit mutual exclusivity. Figure 7.6 shows that this difference increases with larger populations. However, when s_{MAX} becomes many times greater than the number of agents, the disadvantage of switching decreases. That is, agents retain words that have already been discarded by others.

The most rational strategy for any agent is not to assume mutual exclusivity to begin with, and only to activate it under relevant conditions. This reflects the findings that 14-month-old children do not exhibit it while 17-month-olds do (Halberda 2003). From this model we might conclude that mutual exclusivity is an acquired heuristic which is applicable in situations where there is likely to be low variation (monolingualism). More research is required into this kind of model. The point here is that the assumptions of the original model obscure the distinction between mutual exclusivity as an innate, universal bias and an acquired, culture-specific one.

7.5 Discussion

Communicative success can emerge without mutual exclusivity. The results of this model stand in opposition to previous research (e.g. Smith 2002; Vogt and Haasdijk 2010; Hutchins and Hazlehurst 1995; Oliphant 1999 and De Vylder 2006). For instance, it has been claimed that "human language learners appear to bring a one-to-one bias to the acquisition of vocabulary systems. The functionality of human vocabulary may therefore be a consequence of the biases of human language learners" (Smith 2004, p. 127). The current research suggests that mutual exclusivity is not an innate bias. Furthermore, the bias becomes functional as a consequence of the variance in the vocabulary and social dynamic. A related model shows similar results (Smith 2005): Mutual exclusivity is not necessary for communicative success, but helps agents co-ordinate linguistically when they have conceptual differences. Multiple consensus systems can be maintained in a population with complex social structures (de Vylder 2007). However, the current model shows that mutual exclusivity does not always aid the co-ordination process.

However, rather than directly opposing the claims of some previous models, the constraints in the current model can be seen as a relaxation of the constraints embodied by the mutual exclusivity bias. Both models contribute the necessary ingredients for an evolutionary system: Heredity, variation and differential fitness (e.g. Lewontin 1970). Although generational turnover is not modelled, there is heredity in the sense that each agent inherits its own memory from the previous round. Heuristic A introduces the variation by adding new words. Heuristic B introduces differential fitness by selecting words which are successful in communication. Without heuristic A, variation is still introduced by agents creating new words at early stages of the game when they have no words at all (step 4 of the algorithm). The generalisation of Heuristic B to keep an arbitrary number of words after successful communication allows selection to operate over *groups* of words rather than single ones.

Heuristics A and B, then, are an efficient way of introducing the ingredients for evolution into the system. However, cultural processes can also introduce these ingredients – the individual learning processes need not be the source. Other processes could also introduce variation such as errors in production or perception or differences in contact with other agents.

7.6 Conclusion

The naming game was reanalysed in the light of evidence from bilingual language acquisition research. The measurements used to analyse the model were also re-assessed and shown to favour monolingual systems. Steps in the categorisation game were re-analysed as implementing a mutual exclusivity constraint. To explore the effects of these steps, the learning algorithm was generalised so that the steps could

be omitted. Communicative success at the lexical level was achieved without mutual exclusivity constraints. In fact, in some cases, the constraint impedes the process.

This goes against some previous research which argued that mutual exclusivity is necessary for communication to emerge. What seems to be important is the presence of the ingredients for evolution – inheritance, variation and selection. The mutual exclusivity bias is seen as an efficient way of integrating these ingredients. However, the model also showed that rational agents should not assume mutual exclusivity to begin with. This reflects research which shows that children only start using mutual exclusivity in certain situations. Mutual exclusivity is not appropriate in a bilingual environment, so bilinguals do not exhibit it. Given this, the monolingual assumptions of the naming game are unrealistic for two reasons. First, a learner's learning algorithm may change over time, as demonstrated by the differences found between monolinguals and bilinguals. Secondly, they are not valid abstractions because the heuristics which implement mutual exclusivity are optional extras, so the simplest, default assumptions of learners should be bilingual. That is, monolingualism is a specialised form of bilingualism.

When modelling cultural processes, abstraction is necessary. However, the cultural phenomena that appear simplest (e.g. monolingualism) may not be caused by the simplest learning mechanisms. Much of the complexity in cultural phenomena stem from complex interactions *between* individuals. That is, the cultural transmission process itself can shape and influence the cultural practices it transmits.

7.6.1 Integrating Cultures in the Light of Cultural Adaptation

The communication system in the model above adapts to fit the needs and constraints of its users. Indeed, the hallmark of a cultural phenomenon is that it has adapted to the cognitive niche of its community's members (Christiansen and Chater 2008; Beckner et al. 2009). If different communities have different dynamics, such as population size or differences in social structures, then the cultural phenomena that emerge in them may be radically different. In the model above, the communication system between two agents became optimised for efficiency while the communication system in a more complex social structure became optimised for flexibility. Biases in communities towards these different optimisations could be amplified by cultural transmission (Kirby et al. 2007). Over many generations, and for a more complex cultural phenomenon (e.g. a language system, judicial system or musical form), the commonalities between two communities may erode to very abstract principles. When seeking to integrate them, then, a common measure for separate cultures may be difficult to find. Even something as simple as assuming each object only has one associated word may reflect the deep structure of the culture in which it is embedded.

References

Abrams, D.M. and Strogatz, S. H.: Modelling the dynamics of language death. Nature, 424, pp. 900. (2003)

Baker, C. & Jones, S. P.: Encyclopedia of bilingualism and bilingual education. Clevedon, UK: Multilingual Matters. (1998)

Baronchelli, A.: Role of feedback and broadcasting in the naming game. Physical Review E. 83(4), 046103. (2011)

Baronchelli, A., Dall'Asta, L., Barrat, A. & Loreto, V.: Bootstrapping communication in language games in A. Cangelosi, A. D. M. Smith & K. Smith, eds.: The Evolution of Language, Proceedings of the 6th International Conference (EVOLANG6), World Scientific Publishing Company. (2006)

Baronchelli, A., Felici, M. Loreto, V., Caglioti, E. &Luc Steels:Sharp transition towards shared vocabularies in multi-agent systems. Journal of Statistical Mechanics: Theory and Experiment. P06014. (2006)

Baronchelli, A., Gong, T., Puglisi, A., and Loreto, V.: Modeling the emergence of universality in color naming patterns. Proceedings of the National Academy of Sciences, 107(6):2403–2407. (2010)

Burkett, D., Griffiths, T.: Iterated learning of multiple languaged from multiple teachers. In Smith, A., Schouwstra, M., de Boer, B., and Smith, K., editors, The Evolution of Language: Proceedings of EvoLang 2010, 58–65. (2010)

Byers-Heinlein, K. & Werker, J. F.: Monolingual, bilingual, trilingual: infants' language experience influences the development of a word-learning heuristic. Developmental Science, 12(9), 815–823. (2009)

Christiansen, M. H. & Chater, N.: Language as shaped by the brain. Behavioral and Brain Sciences, 31, 5, 489–508; discussion 509–58. (2008)

De Beule, J. and De Vylder, B. and Belpaeme, T. A cross-situational learning algorithm for damping homonymy in the guessing game. In L. M. Rocha, L. Yaeger, M. Bedau, D. Floreano, R. Goldstone, and A. Vespignani (eds.) Artificial Life X, 466–472. MIT Press. (2006)

De Vylder, B. & Tuyls, K.: How to reach linguistic consensus: A proof of convergence for the naming game. Journal of Theoretical Biology, 242 (4) 818–831. (2006)

de Vylder, B. The Evolution of Conventions in Multi-Agent Systems. PhD thesis, Artificial Intelligence Laboratory, Vrije Universiteit Brussels. (2007)

Euopean Comission. 2006. Special Eurobarometer: Europeans and their Languages: European Commission. http://ec.europa.eu/public_opinion/archives/ebs/ebs_243_en.pdf.cited3Nov2010.

Gong, T., Puglisi, A., Loreto, V., & Wang, W. S-Y.: Conventionalization of linguistic knowledge under simple communicative constraints. Biological Theory: Integrating Development, Evolution, and Cognition, 3 (2) 154–163. (2008)

Griffiths, T. L. & Kalish, M. L.: Language Evolution by Iterated Learning With Bayesian Agents. Cognitive Science: A Multidisciplinary Journal. 31 (3) 441–480. (2007)

Halberda, J.: The development of a word-learning strategy. Cognition, 87 (1) 23–34. (2003)

Healey, E. & Skarabela, B.: Are children willing to accept two labels for one object? A comparative study of mutual exclusivity in bilingual and monolingual children. In Proceedings of the Child Language Seminar. University of Reading. (2009)

Houston-Price, C., Caloghiris, Z., & Raviglione, E.: Language Experience Shapes the Development of the Mutual Exclusivity Bias. Infancy 15 (2), 125–150. (2010)

Hutchins, E., & Hazlehurst, B.: How to invent a lexicon: The development of shared symbols in interaction. In N. Gilbert & R. Conte (eds.) Artificial societies: The computer simulation of social life, pp. 157–189, London: UCL Press. (1995)

Lewontin, R. C.: The Units of Selection. Annual Reviews of Ecology and Systematics 1: 1–18. (1970)

Loreto, V., Baronchelli, A., and Puglisi, A.: Mathematical Modeling of Language Games, Evolution of Communication and Language in Embodied Agents, part 3, 263–281. Springer Berlin Heidelberg. (2010)

Luchtenberg, S.: Bilingualism and bilingual education and their relationship to citizenship from a comparative German-Australian perspective, Intercultural Education, Vol. 13(1), 49–61. (2002)

Kirby, S.: Spontaneous evolution of linguistic structure-an iterated learning model of the emergence of regularity and irregularity. IEEE Transactions on Evolutionary Computation, 5 (2) 102–110. (2001)

Kirby, S., Dowman, M. & Griffiths, T.:Innateness and culture in the evolution of language. Proceedings of the National Academy of Sciences, 104:12, 5241–5245. (2007)

Komarova, N. L. & Jameson, K.A. Population heterogeneity and color stimulus heterogeneity in agent-based color categorization.Journal of Theoretical Biology. 253(4): 680–700. (2008)

Kostoulas-Makrakis, N.: Language and Society. Basic Concepts (in Greek). Athens: Metexmio. (2001)

Markman, E. M. and Wachtel, G. A.: Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology*. 20, 121–57. (1988)

Merriman, W. E. and Bowman, L. L.: The mutual exclusivity bias in children's word learning. *Monographs of the Society for Research in Child Development*. 54 (3–4). (1989)

Niyogi, P. & Berwick, R. C.: The Logical Problem of Language Change. CBCL Paper 115, MIT AI Laboratory and Center for Biological and Computational Learning, Department of Brain and Cognitive Sciences. (1995)

Niyogi, P. and Berwick, R.C.: The proper treatment of language acquisition and change in a population setting. Proceedings of the National Academy of Sciences. 106 (25) 10124–9. (2009)

Nowak, M. A., & Krakauer, D. C.: The evolution of language. *Proceedings of the National Academy of Sciences of the United State of America*. 96 (14) 8028–8033. (1999)

Oliphant, M.: The learning barrier: Moving from innate to learned systems of communication. Adaptive Behavior, 7, 371–384. (1999)

Pearson, B.Z., Fernandez, S.C., Oller, D.K.: Lexical development in bilingual infants and toddlers: Comparison to monolingual norms. Language Learning, 43, 93–120. (1993)

Puglisi, A., Baronchelli, A. & Loreto, V.: Cultural route to the emergence of linguistic categories *Proceedings of the National Academy of Science*. 105 7936–7940. (2008)

Smith, K: The cultural evolution of communication in a population of neural networks. Connection Science, 14, 65–84. (2002)

Smith, K.: The evolution of vocabulary. Journal of Theoretical Biology. 228 (1) 127–142. (2004)

Smith, A. D. M.: Mutual exclusivity: communicative success despite conceptual divergence. In M. Tallerman (ed.) Language Origins: perspectives on evolution. 372–388. Oxford University Press. (2005)

Smith, K.: Iterated learning in populations of Bayesian agents. Proceedings of the 31st Annual Conference of the Cognitive Science Society. (2009)

Smith, K., Kirby, S. & Brighton, H.: Iterated learning: a framework for the emergence of language. Artificial Life 9 (4) 371–86. (2003)

Statistics Canada: The Evolving Linguistic Portrait, 2006 Census Statistics Canada Catalogue no. 97–555-XIE. Ottawa. Version updated May 2010. http://www12.statcan.ca/census-recensement/2006/as-sa/97-555/pdf/97-555-XIE2006001-eng.pdf (3/11/2010). (2007)

Steels, L.: Perceptually grounded meaning creation in M. Tokoro (Ed.) *Proceedings of the International Conference on Multi-agent Systems*. MIT Press. Cambridge, MA. (1996)

Steels, L. and Belpaeme, T.: Coordinating perceptually grounded categories through language: a case study for colour. Behavioral and Brain Sciences, 28(4):469–489. (2005)

The "Five Graces Group": Beckner, C., Blythe, R., Bybee, J., Christiansen, M. H., Croft, W., Ellis, N. C., Holland, J., Ke, J., Larsen-Freeman, D. & Schoenemann, T. Language Is a Complex Adaptive System: Position Paper. Language Learning, 59, 1–26. (2009)

Thomas, L., & Wareing, S.: Language, Society and Power. London: Routledge. (1999)

U.S. Census Bureau: Language Use and English-Speaking Ability: 2000 (C2KBR-29). Washington: Government Printing Office. http://www.census.gov/prod/2003pubs/c2kbr-29.pdf (3/11/2010). (2003)

Vogt, P. & Haasdijk, E.:Modeling social learning of language and skills. Artificial Life 16 (4) 289–309. (2010)