Language Learning and Language Contact

Luc Steels

VUB Artificial Intelligence Laboratory
Pleinlaan 2 1050 Brussels
email: steels@arti.vub.ac.be and
Sony Computer Science Laboratory
6 Rue Amyot
F-75005 Paris, France

The evolution of language can only be explained when we take a language learning process into account which is necessarily imperfect due to weak data and the limits of induction. Thus language change yields clues on what the learning processes are and conversely hypothesised learning processes should predict possible language changes. This paper considers this issue by studying lexicon formation processes. It is shown how a process for the construction and acquisition of a lexicon in a single agent leads to various dynamical phenomena when applied to a population of agents with varying degrees of contact.

1 Introduction

It is a fact that natural languages evolve at all levels: New sounds may get introduced or the sound system of a language may start shifting, new words and meanings may enter in the lexicon or existing meanings may expand or contract, new syntactic categories may emerge, syntactic structures may shift, features expressed through morphology may shift to a syntactic expression, etc. Sometimes these changes are very rapid, particularly when a new 'creole' language is formed [3]. There may also be periods of relative stability. All these changes take place against a background of continuously changing populations (the linguistic community is in constant flux due to the in- and outflow of members) and changing environments and thus new sources of meaning (new artefacts, new institutions, new types of social relations, etc.).

As long as we restrict linguistic investigations to a single (idealised) speaker-hearer in a homogeneous language community who knows and speaks the language perfectly [1], it is a puzzle why languages change. It is even more a puzzle if we assume that language is to a large extent innate, which implies that true innovations can only occur through genetic mutations and subsequent spreading in the population [4]. Genetic mutations are simply too slow compared to the time-scales observed in the emergence and spreading of linguistic innovations, and success of a linguistic innovation depends on whether the group has adopted it, i.e. a mutation in a single agent does not give this agent any immediate advantage.

On the other hand, if we accept that languages are invented and transmitted culturally, language change becomes self-evident. During transmission, and particularly during learning, variations will automatically arise and these variations may become conventionalised as they propagate to the rest of the group. Language learning, which is at the heart of the transmission process, then provides us important clues about what kind of languages changes are to be expected and, conversely, changes in languages give us clues on what kind of learning processes must be active. For example, language learners (as listeners) may over-interpret linguistic behavior, in the sense that they may believe that certain behavior is rule-governed whereas it is not. If the over-interpretation is however then used by the learners in subsequent language production, it can become a new norm. An empirical study of the kind of over-interpretation that has taken place, may give us a clue on what kind of learning strategies were at work.

The cause of language changes are partly internal [2]. For example, if a new sound starts to propagate, it may cause a chain shift of changes in existing sounds so that distinctiveness is retained in the total system. Other causes of language change are external. They come from contact between languages due to increased communication between speaker communities, large-scale bilingualism causing mutual influence, the splitting of a community in separated groups causing subsequent language divergence, etc.

This paper focuses on language change due to changes in the dynamics of the underlying populations, particularly the communication structure between linguistic communities. A very basic linguistic behavior is investigated: the naming of objects with the aim of developing a common lexicon. The learning/construction procedure is kept as simple as possible so as to focus on the issue of language contact itself. Even with these conditions, we see interesting phenomena emerge which cannot be explained by focusing on single speakers without learning capability.

The rest of the paper is in three parts. The next part introduces the linguistic behavior and the learning process of a single agent. Then the typical behavior of such a system is studied without changes in language contact. Finally an experiment is discussed related to the dynamics of language contact.

2 The naming game

2.1 Definition

The *naming game*, as introduced by Steels [7], is a specific kind of language game which is played between two agents. It is a theoretical model for studying how a shared vocabulary may emerge in a distributed group of agents based only on local interaction. One agent (the speaker) identifies an object by using a name. The game succeeds if the other agent (the hearer) agrees that this name is appropriate for the object identified. A naming game is adaptive when both participants in the game change their rules (i.e. their word-object associations) to be more successful in future games. The naming game has been investigated through computer simulations as well as in robotic experiments in which the meanings are grounded in the sensori-motor behavior of the robots [9].

More formally, there is a set of *agents* $\mathcal{A}$ of size $\mathcal{N}_{\mathcal{A}}$ where each agent $a \in \mathcal{A}$ has contact with a set of *objects* of size $\mathcal{N}_{\mathcal{O}}$, $\mathcal{O}_a = \{o_0, ..., o_n\}$. In the present paper it is assumed that the set of agents and the set of objects is identical, i.e. the agents talk only about themselves. A *word* is a sequence of letters drawn from a finite alphabet. The agents are all assumed to share the same alphabet. A *lexicon* $\mathcal{L}$ is a relation between objects and words. Each member of the relation has two associated quantities: The number of times the relation has been used and the number of times the relation was successful in use. Each agent $a \in A$ has his own lexicon $L_a \subset \mathcal{O}_a \times W_a \times \mathcal{N} \times \mathcal{N}$ which is initially empty. There is the possibility of synonymy and homonymy: A single word can be associated with several objects and a given object can be associated with several words. An agent $a \in \mathcal{A}$ can now be defined as a pair $a =< \mathcal{L}_a, \mathcal{O}_a >$.

A *naming game* $N =< s, h, o >$ is an interaction between two agents: a speaker s and a hearer h about a topic o which is an object, $o \in \mathcal{O}_s$. The interaction proceeds as follows:

1. The speaker randomly selects a topic out of his set of objects. The speaker draws the attention of the hearer to the topic. In a natural conversation this could be by pointing or prior linguistic exchanges.
2. The speaker s encodes the topic o through a word w. The word chosen is the most successful word w where $< o, w, u, s > \in \mathcal{L}_s$. A word w_1 is more successful than a word w_2 iff $< o, w_1, u_1, s_1 > \in \mathcal{L}_s$, $< o, w_2, u_2, s_2 > \in \mathcal{L}_s$, $u_1 \geq u_2 \geq 0$ and either $s_1/u_1 > s_2/u_2$ or $s_1/u_1 = s_2/u_2$ and $u_1 > u_2$. If several words are equally successful, a word is selected randomly from the most successful set.
3. The hearer h decodes the word w to derive a set of objects H such that $h \in H \Longrightarrow < h, w, u, s > \in \mathcal{L}_h$
4. A naming game is successful iff $o \in H$.

Several naming games involving different agents can be going on in parallel because interactions are always local.

An *adaptive naming game* is a naming game $N =< s, h, o >$ as defined above, with changes to both the speaker and hearer as a side effect of the game. This means that we have to introduce a temporal dimension in the definition of an agent. An agent a at time t is defined as $a_t =< \mathcal{L}_{a,t+1}, \mathcal{O}_{a,t} >$. A time point corresponds to an event in which two agents play a language game. If the speaker uses a word w where $< o, w, u, s > \in \mathcal{L}_{s,t}$ and if the game ends in success, then $< o, w, u + 1, s + 1 > \in \mathcal{L}_{s,t+1}$ and $< o, w, u + 1, s + 1 > \in \mathcal{L}_{h,t+1}$. If a word is used by speaker or hearer but some failure has occured, then use is incremented but not success: $< o, w, u + 1, s > \in \mathcal{L}_{s,t}$ and $< o, w, u + 1, s > \in \mathcal{L}_{h,t+1}$.

The other changes defining the state of an agent at time $t + 1$ are as follows:

1. Missing object(step 1 fails)

The object chosen as topic by the speaker is not shared by the hearer: $o \notin \mathcal{O}_{h,t}$. In that case, the game ends in failure but the hearer may acquire the object with an object propagation probability p_o $\mathcal{O}_{h,t+1} = \mathcal{O}_{h,t} \bigcup \{o\}$.

2. The speaker does not have a word.(step 2 fails)

In other words there is no $< o, w, u, s > \in \mathcal{L}_{s,t}$. In that case, the game fails but the speaker may (randomly) create a new word w' and associate this with the object in his lexicon with a word creation probability p_c, $\mathcal{L}_{s,t+1} = \mathcal{L}_{s,t} \bigcup \{< o, w', 1, 0 >\}$.

3. The hearer does not know the word. (step 3 fails)

In other words there is no $< o, w, u, s > \in \mathcal{L}_{h,t}$. In that case, the game ends in failure but the hearer may extend his lexicon with a word absorption probability p_a, $\mathcal{L}_{h,t+1} = \mathcal{L}_{h,t} \bigcup \{< o, w, 1, 0 >\}$.

4. The topic is not decoded by the hearer. (step 4 fails)

In other words, $o \notin H$. In that case the game ends in failure, but the hearer may acquire the word for his lexicon with a word absorption probability p_a: $\mathcal{L}_{h,t+1} = \mathcal{L}_{h,t} \bigcup \{< o, w, 1, 0 >\}$

If there is no explicit change as defined above, then the definition of an agent a remains the same: $a_{t+1} =< \mathcal{L}_{a,t}, \mathcal{O}_{a,t} >$

Note that these rules imply both a learning activity (steps 3,4) and a construction activity (step 2). Indeed, the approach here is not purely inductive but constructive - which is necessary if we want to explain how new lexicons may originate from scratch or how new lexicalisations may ever take place. As a consequence of this constructive step, lexicons become open: At any time new meanings (which in this case means new agents) may be introduced and the lexicon expands automatically.

2.2 Global properties of lexicons

As will become clear later, a population of agents only becomes completely coherent when the population is closed and left to evolve to reach coherence for a long time. Clearly these are unususal (and uninteresting) circumstances. If the system is open, every agent develops its own ideolect which is more or less similar to that of other agents in the population but never quite the same. The following measures are relevant for characterising global properties of these lexicons:

1. The *communicative success* is the outcome of a language game. This success is 1.0 if the game succeeds and 0.0 if it does not, for whatever reason. $S_{\mathcal{A},k}$ is defined as the average communicative success of a set of k language games in a population $\mathcal{A}$. It is an indication in how far the agents in the community succeed in communication.

2. The *coherence* $C_{\mathcal{A},t}$ of a lexicon in a group of agents $\mathcal{A}$ measures the average percentage of the most preferred words in the lexicon. It is an indication how strong the agents have converged on prefered word-meaning associations.

3. Because agents may maintain more than one word for the same meaning, they may understand a word (and thus reach communicative success) even if they would themselves prefer another word. The *mutual understanding* measure characterises how much of an agent's prefered words are known by another agent, without the other agent necessarily prefering the same word. This measure is generalised to an *intercluster understanding* measure $U_{\mathcal{A}_1,\mathcal{A}_2,t}$ which quantifies the average mutual understanding between two groups of agents $\mathcal{A}_1$, $\mathcal{A}_2$ at time t. This measure can be interpreted as quantifying bilingualism, because the agents

of cluster $\mathcal{A}_1$ know the words used by the other cluster but do not use them themselves, and vice-versa.

4. The *mutual equality* $E_{\mathcal{A}_1,\mathcal{A}_2,t}$ is equal to the average percentage of shared word-meaning pairs between two groups of agents $\mathcal{A}_1$ and $\mathcal{A}_2$ at time t. A word-meaning pair is shared if both agents have adopted it as the most prefered word for expressing a particular meaning.

3 Behavior of Multi-Agent Naming Game Systems

In an earlier paper [7], the behavior of groups of agents engaged in naming games has been studied. In closed naming game systems, i.e. where the set of agents and possible meanings is constant, total communicative success is quickly reached (figure 1).

Figure 1. Simulation of closed system with 20 agents and thus 20 possible meanings. Communicative success is shown on the y-axis per 25 games. A rapid increase in communicative success is achieved for the first 200 x 25 games, after which progress levels off to reach complete average success after 400 x 25 games.

An inspection of the lexicon of the invididual agents shows that many more words are 'floating around' in the population than those that finally become the norm. Here is part of the lexicon of agent a255. For each object we show the possible words followed by the use, success and score. The words are ordered according to the success criterion used by the agent to decide which word to prefer.

```
Obj255:  BIPA[34,34,1.00] DUPO[1,1,0.00]
        NUJO[0,0,0.00]
```

```
Obj254:  BEGAGO[22,22,1.00] GU[29,21,0.72]
        DE[2,2,0.00] BADO[1,1,0.00]
        NEKIKO[0,0,0.00] MIPA[0,0,0.00]
        GUGIGE[0,0,0.00]
Obj253:  KUJO[33,33,1.00] TUGAJA[10,5,0.50]
        TI[0,0,0.00] BU[0,0,0.00]
Obj252:  TEPI[21,21,1.00] GUDEBA[8,4,0.50]
        DU[4,4,0.00] GIMO[1,1,0.00]
        JOJEJI[0,0,0.00]
Obj251:  DEGUPI[3,3,0.00]
 ...
```

We observe that only a few objects have a single word (one of them is Obj251). For all others there is synonymy: Different words are used to denote the same object, although there is typically one word that dominates and many words have no success at all and could easily be removed. Domination of one word will reenforce itself as games proceed (figure 2).

Figure 2. This figure shows the different competing words ("bipa", "nujo", "dupo") in the population for the same object. After a certain set of games, there is a convergence to a single word ("bipa").

Together with Angus McIntyre, naming games have also been investigated as open systems [?], by imposing an in- and outflux of agents. Because the set of possible meanings is equal to the set of agents, this automatically means that the set of meanings also changes. Figure 3 shows what happens with low birth and death rates. Even though the members of the population change and therefore also the objects they talk about, there is no loss in communicative success which means that the population is able to cope both with in- and outflux of agents and in- and outflux of objects. Note that the lexicon will change as time progresses. After a while the population will be entirely renewed and the lexicon will be completely different from the initial situation because the set of possible meanings is entirely different.

Figure 3. The figure shows the evolution of the communicative success in a fluctuating population of agents. New agents enter with a probability 0.00005 and the depart with the same probability. The population is able to cope with the flux.

Steels and McIntyre have also investigated what happens when a spatial structure is introduced [?]. Each agent a has x,y coordinates a_x, a_y. The probability with which two agents a and b interact is based on their respective trigonometric distance, and an interaction factor m_a, which determines the weight of the distance. A spatial distribution is randomly generated, modulated by a clustering tendency. A typical example of a distribution is given in figure 4.

A typical result of lexicon formation with fairly weak interaction is shown in figure 5. The agents develop a stable lexicon within their cluster, but they will at the same time become familiar with words in other clusters. In effect, they become bilingual. The following lexicons illustrate this clearly. The first one is taken from an agent from the left most cluster in figure 3:

```
...
obj3411:  MU[49,48,0.98] GITE[4,3,0.00]
        PAKU[2,1,0.00]
obj3410:  NA[16,15,0.94] NUGUGE[21,19,0.90]
        PA[5,4,0.80] NE[3,0,0.00]
```

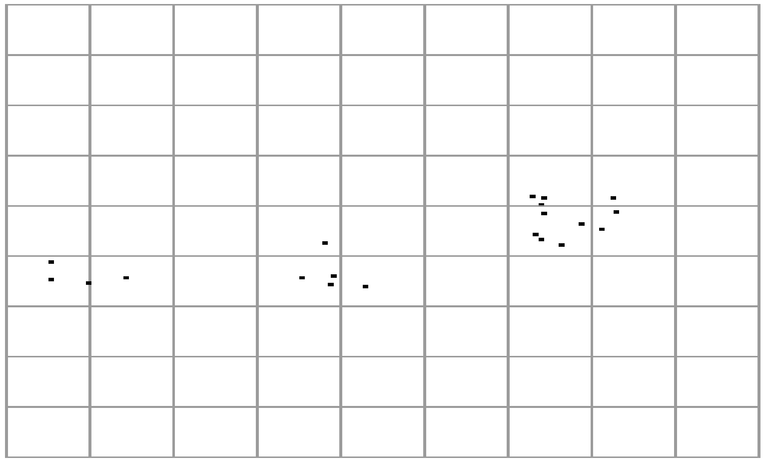

Figure 4. The figure shows the spatial distribution of a set of 20 agents. There is clustering around three centers.

```
obj3409:  PE[51,48,0.94] DA[2,1,0.00]
obj3408:  KUDE[48,43,0.90] NADO[3,2,0.00]
obj3407:  MEPABO[38,37,0.97] JABETO[17,12,0.71]
        DI[1,0,0.00]
obj3406:  BO[35,33,0.94] BABIGE[11,9,0.82]
obj3405:  TINE[47,46,0.98]
obj3404:  NEBU[42,41,0.98] ME[12,10,0.83]
obj3403:  MOMA[44,43,0.98] TI[3,0,0.00]
        NUDO[3,0,0.00] KENE[3,2,0.00]
obj3402:  GO[59,58,0.98] TA[1,0,0.00]
obj3401:  NUGINI[31,30,0.97] GI[18,15,0.83]
        MAJIBA[4,0,0.00]
obj3400:  KUBE[43,38,0.88] GUTIDA[4,3,0.00]
        MOKO[1,0,0.00]
```

The second is from an agent from the right most cluster:

```
...
obj3411:  GITE[42,37,0.88]
obj3410:  PA[40,35,0.87]
obj3409:  DA[37,32,0.86] NINE[7,5,0.71]
obj3408:  NADO[24,23,0.96] KUDE[10,8,0.80]
        PUTO[1,0,0.00]
obj3407:  MEPABO[40,36,0.90] JUNIPE[8,6,0.75]
        DI[1,0,0.00]
obj3406:  BABIGE[42,39,0.93] BO[3,2,0.00]
obj3405:  TINE[71,64,0.90]
obj3404:  NEBU[38,37,0.97] ME[1,0,0.00]
```

```
        BUKUGO[1,0,0.00]
obj3403:  KENE[27,24,0.89] MOMA[17,14,0.82]
        NUDO[4,0,0.00] KOKO[1,0,0.00]
obj3402:  GO[19,18,0.95] TA[5,1,0.20]
obj3401:  GI[9,8,0.89] MATU[20,17,0.85]
        PUMONI[7,0,0.00]
obj3400:  GUTIDA[33,26,0.79] KUBE[4,3,0.00]
```

Some words (for example "TINE" for obj3405) are shared. Sometimes there are no shared words (yet). For other objects there are at least two words. One used preferentially inside the cluster, the other known but preferentially used by members of another cluster. Thus the word "KUBE" for obj3400 is preferentialy used by the first agent whereas the word "GUTIDA", although known by the first agent is prefered by the second one.

Figure 5. The figure shows the evolution of communicative success in a group of 20 agents clustered in three communities. Total communicative success is slower to be reached because there are spurious interactions with members of the other community.

4 The Dynamics of Language Contact

This brings us to the subject of the present paper. Given a spatial distribution and a variable degree of interaction, we can study the impact of increased communication and hence more intense language contact on clusters of language communities.

The experiment runs as follows. Agents in three different clusters (a, b, and c) interact and evolve a lexicon for each cluster. The interaction between clusters is initially very weak (see figure 6). At some point in time the intercluster communication is increased drastically. At first there is a drop in communicative success but then total communicative success is again reached.

Figure 6. Evolution of average communicative success in a Group of agents with first weak and then strong interactions. (the change maps onto point 122). The average communicative success per 25 games is shown at each point.

This general evolution hides however the more interesting developments. Figure 7 shows the evolution of coherence $C_{\mathcal{A},t}$ for each cluster (a, b, c) separately and also for the total set of agents. As long as the agents have relatively little contact, total coherence is low although coherence of each cluster is high. Coherence starts to increase with increased contact. Coherence in each cluster diminishes somewhat because the agents in the cluster are in the process of accomodating to the total set.

The next figure (figure 8) shows with the same scale the evolution of the equality measure $E_{\mathcal{A}_1,\mathcal{A}_2,t}$ between clusters a and b, a and c, and b and c. We see that the equality reaches a plateau before the increase in contact but then starts to climb. This means that the languages of the different groups are in the process of merging due to the increased language contact.

Figure 7. The figure shows the evolution of coherence plotted for each 100 games. We see that initially each individual cluster reaches almost total coherence, whereas the total set of agents has a much lower coherence. After increased contact (which maps onto point 17), we see that there is a steady increase in total coherence.

Finally, mutual understanding or bilingualism is shown in the next figure (figure 9). The moment language contact increases we see a rapid increase in understanding, i.e. in how far a prefered word of one agent is also known to the other agent.

Thus we see that increased contact causes at first a rapid increase in bilingualism, then a gradual mixing of the languages, and, if the contact continues, an evolution towards complete coherence. The more rapid the contact is increased, the faster the three phases can be observed. Interestingly enough the communicative success of the whole system (intra-cluster and inter-cluster communication) is hardly affected, as shown in figure 6.

These various phases are also seen in natural languages [10]. In some areas, such as in central Belgium, two language communities (French and Flemish) co-exist and large groups of the population have become bilingual. In these cases we see a strong mutual influence on the lexicons of both communities. Thus a cup will be called "kop" in Dutch but "tas" in Flemish, influenced by the French "tasse". The homonym "tas" means bag in Dutch as well as Flemish. In other cases, a subpopulation that has been exposed due to invasion or migration to a larger more dominating language community may progressively or radically shift their lexicon to the language of the majority group, even though the grammar remains intact. Many such cases exist for pidgin and creole languages. Finally, lexicon convergences are currently seen in dialectal language groups (for example

Figure 8. The figure shows the evolution of how many prefered words are shared between groups of agents. This starts to increase slowly after language contact has increased (around point 17).

in the different dialects of French) due to increased language contact.

Obviously a lot more experiments can be done:

1. Population sizes could be varied to see in how far population domination leads also to language domination.
2. Other factors could be made to influence the probability of inter-cluster communication. For example, the social distance could have the same stratification effect as spatial distance.
3. Rather than increasing language contact, we could decrease it, showing that the population diversifies.
4. The populations in the cluster experiment above are closed. Another obvious experiment is to let them vary at specific rates, so that the languages in each cluster are moving targets.

All these experiments reflect situations that occur in human language populations and can thus be compared to language evolution phenomena that have been observed to accompany population changes.

Figure 9. Evolution of intercluster understanding or bilingualism. A rapid increase is seen after agent contact increases which happens around point 17.

5 Conclusion

Languages are known to evolve when the community comes into contact with the language of another community. This evolution can be towards bilingualism, mixing, and eventually a total coherence in which the differences between the languages disappear. The key point of the present paper is that these phenomena can only be understood when we assume that language speakers invent and transmit their language in a cultural fashion.

This was demonstrated by the investigation of a very basic linguistic function, namely naming, and a learning/construction procedure whereby agents invent and transmit names culturally. Nothing had to be added or changed to the linguistic behavior of single agents in order to get the dynamics of divergence and coherence known to happen in natural language contact. Similar experiments could be done for other linguistic capabilities, such as the development and maintenance of a shared phonology, or the development and maintenance of a shared set of syntactic conventions (a grammar). In this case, the games will be of a different nature and the accompanying adaptation and construction processes will be much more complex. But the basic structure of the dynamics caused by variations in contact will be similar.

An additional consequence of the results presented in this paper is that language convergence does not necessarily have to be caused by a common origin.

The fact that languages share vocabularies, as investigated in statistically oriented research towards cognates (shared words) [6], can be the result of interaction between clusters of language users, and is not necessarily the result of a universal common *Ursprache* as comparative linguists often assume.

6 Acknowledgement

This work was financed and carried out at the Sony Computer Science Laboratory in Paris. We are indebted to Toshi Doi and Mario Tokoro to be able to work in this superbly productive environment. The BABEL software tool under development by Angus McIntyre at SONY CSL-Paris has proved invaluable for efficiently exploring the issues discussed in this paper.

References

1. Chomsky, N. (1975) Reflections on Language. Pantheon, New York.
2. Labov, W. (1994) Principles of Linguistic Change. Volume 1: Internal Factors. Blackwell, Oxford.
3. Arends. J, P. Muysken and N. Smith (1995) Pidgins and creoles: an introduction. Benjamins, Amsterdam.
4. Pinker, S. (1994) The language instinct. Penguin Books. London.
5. Renfrew, C. (1987) Archeology and Language. The Puzzle of Indo-European Origins. French Translation: Champs Flammarion, Paris.
6. Ruhlen, M. (1994) On the Origin of Languages. Studies in Linguistic Taxonomy. Stanford University Press, Stanford Ca.
7. Steels, L. (1996) Self-organizing Vocabularies. In: Langton, C. (ed.) (1996) Proceedings of the V Alife Conference, Nara Japan.
8. Steels, L. and A. McIntyre (1997) Spatially Distributed Naming Games. Submitted to ECAL-97, Brighton.
9. Steels, L. and P. Vogt (1997) Grounding adaptive language games in robotic agents. Submitted to ECAL-97, Brighton.
10. Thomason, S. and T. Kaufman (1988) Language Contact, Creolization, and Genetic Linguistics. University of California Press, Berkeley.