The Recruitment Theory of Language Origins

Luc Steels
SONY Computer Science Laboratory - Paris.
AI Laboratory Vrije Universiteit Brussel

No Institute Given

Abstract. The recruitment theory of language origins argues that language users recruit and try out different strategies for solving the task of communication and retain those that maximise communicative success and cognitive economy. Each strategy requires specific cognitive neural mechanisms, which in themselves serve a wide range of purposes and therefore may have evolved or could be learned independently of language. The application of a strategy has an impact on the properties of the emergent language and this fixates the use of the strategy in the population. Although neurological evidence can be used to show that certain cognitive neural mechanisms are common to linguistic and non-linguistic tasks, this only shows that recruitment has happened, not why. To show the latter, we need models demonstrating that the recruitment of a particular strategy and hence the mechanisms to carry out this strategy lead to a better communication system. This paper gives concrete examples how such models can be built and shows the kinds of results that can be expected from them.

Keywords: Language origins, language evolution, language games, recruitment theory.

1 Introduction

Tremendous progress has been made recently on the fascinating question of the origins and evolution of language (see e.g. [54], [6], [8], [30]). There is no widely accepted complete theory yet, but several proposals are on the table and observations and experiments are proceeding. This article focuses on the recruitment theory of language origins which we have been exploring for almost ten years now. The theory is introduced (in section 3) against the background of the language as adaptation theory (section 2). Next possible methodologies for testing the recruitment theory are discussed (section 4) and three examples are given of computational and robotic experiments (sections 5-7). The article concludes with a discussion of issues for further research.

2 Language as an Adaptation

Recently Pinker [32] and Jackendoff [33] reiterated in their characteristically lucid style the theory that the human language faculty is a complex adaptation that evolved by natural selection for communication ([32], p.16), in other words, that the interconnected areas of the brain involved in language, form a highly specialised neural subsystem, a kind of organ, which is genetically determined and came into existence through Darwinian genetic evolution. Pinker contrasts this language-as-adaptation hypothesis with two alternatives: theories arguing that language is a manifestation of more general cognitive abilities which are in themselves adaptations (ascribed to Tomasello [50], a.o.), and theories arguing that although there is a genetic basis for the language faculty, it has evolved by mechanisms other than natural selection (ascribed to Chomsky [9], a.o.).

Pinker surveys three types of arguments in favor of the language-as-adaptation theory. Arguments of the first type are based on examining the structure of language. Human languages exhibit a number of non-trivial universal trends which could be a logical consequence of an innate language acquisition device [9], particularly because it is difficult to see how the intricate complexity of human language is so easily and routinely acquired by very young children based on apparently poor data [51], and because these same universal trends show up when new languages form, as in the case of creoles [2].

The second type of arguments surveyed by Pinker comes from molecular and population genetics. They rest on the identification of genes that are unique for the human lineage, have undergone selection, and have a clearly identified effect on language [48]. The most prominent example in this respect is the FOXP2 gene which is linked to a specific constellation of language impairments identified in a particular multi-generational family [13]. The third type of arguments in favor of the language-as-adaptation hypothesis is more recent and comes from mathematical and computational investigations which show how a stable communication system might evolve from repeated pairwise interactions and, crucially, whether such systems have the major design features of human language[[32], p. 16]. Computer simulations of the genetic evolution of communication systems ([20], [7], [6]) have indeed shown that certain features of human language, such as bi-directional (or Saussurean) use of signs, can emerge in evolutionary processes, if the competence for communicative success has a direct impact on fitness. This research is complemented by mathematical arguments based on evolutionary game theory, which attempt to show why compositionality or the expression of predicate-argument structure might have become part of the human phenotype if they functionally improve communication [31].

3 The Recruitment Theory

In this article, I introduce and discuss an alternative to the language-as-adaptation hypothesis: the recruitment theory of language origins. This theory hypothesises that the human language faculty is a dynamic configuration of brain mechanisms, which grows and adapts, like an organism, recruiting available cognitive/neural resources for optimally achieving the task of communication, i.e. for maximising expressive power and communicative success while minimising cognitive effort in terms of processing and memory. The implied mechanisms are not specific for language and they are configured dynamically by each individual, and hence genetic evolution by natural selection is not seen as the causal force that explains the origins of language.

One example of recruitment, to be discussed later in more detail, concerns egocentric perspective transformation (computing what the world looks like from another viewpoint). This activity is normally carried out in the parietal-temporal- occipital junction [55] and used for a wide variety of non-linguistic tasks, such as prediction of the behavior of others or navigation [22]. All human languages have ways to change and mark perspective (as in your left versus my left), which is only possible if speaker and hearer can conceptualize the scene from the others perspective, i.e. if they have recruited egocentric perspective transformation as part of their language system. Another example of a universal feature of human languages is that the emotional state of the speaker can be expressed by modulating the speech signal. For example, in case of anger, the speaker may increase rhythm and volume, use a higher pitch, a more agitated intonation pattern, etc. This requires that the neural subsystems involved in emotion (such as the amygdala) are somehow linked into the language system so that information on emotional states can influence speech production and that information from speech recognition can flow towards the brain areas involved with emotion. The recruitment theory argues that recruitment of these various brain functionalities is epigenetic, driven by the need to build a better, more expressive communication system as opposed to genetically pre-determined and evolved through biological selection.

Given that there are many conflicting constraints operating on language (e.g. less effort for the hearer may imply more effort for the speaker) and historical contingencies, we cannot expect that language users will ever arrive at an optimal communication system. On the contrary, there is overwhelming evidence from the historical development of all human languages that language users move around in the search space of possible solutions, sometimes optimising one aspect (for example dropping a complex case system) which then forces another solution with its own inconvenience (for example using a large number of verbal patterns with idiomatic prepositions - as in English). Consequently, language itself can be viewed as a complex adaptive system [39] that adapts to exploit the available physiological and cognitive resources of its community of users in order to handle their communicative challenges, but without ever reaching a stable state.

The recruitment theory resonates with several other proposals for the evolution of language. For example, the biologist Szathmary has put forward the metaphor of a growing language amoeba, a pattern of neural activity that is essential for processing linguistic information and grows in the habitat of a developing human brain with its characteristic connectivity pattern [49]. Many researchers, including those who believe that some parts of the language faculty are innate, agree that a multitude of non-linguistic brain functions gets recruited for language - if we take the language faculty in a broad sense, including conceptualization of what to say [16]. And even Pinker, Jackendoff, et al., who argue for the innateness of many aspects of language, recognise that, at least initially, many of these mechanisms, such as the use of the vocal tract for speech or associative memories to store the lexicon, are pre-adaptations which have been recruited and then become genetically innate through a Baldwinian process of genetic assimilation with possibly further extensions or alterations under the selectionist pressure of language [33], [6].

4 Methodology

In order to explore the recruitment theory, we can use several methods, analogous to the ones reviewed by Pinker [32]:

(1)We can study what universal trends appear in human languages. However rather than simply assuming that they are imposed by the innate language organ, we now try to show that they are emergent properties of solving the communication task, given the constraints of human embodiment, available cognitive mechanisms, properties of the real world, and the inherent difficulties of communication beween autonomous agents who do not have direct access to each others mental states and need to interact about things in the real world as experienced through their sensori-motor system. If some properties of language are emergent in this sense, then they do not have to be genetically coded because they will spontaneously arise whenever human beings engage in building a communication system.

A particularly nice demonstration of this has come from recent investigations into the origins of vowel systems. It is well known that there are clear universal trends in human vowel systems [35] and those adopting the adaptationist view have argued that the features of the vocal tract that enable them, such as the lowering of the larynx [26], or even the distribution of vowels and vowel boundaries themselves [25], have genetically evolved under selectionist pressure for language. On the other hand, it has been shown that those same universal trends reflect constraints of human embodiment (both of the articulatory system and the auditory system) as well as optimality constraints on sound recognition and sound reproduction [27]. Moreover recent simulations carried out in my group [12],[34] have shown that agents with similar embodiment constraints as human beings are able to autonomously self-organise vowel systems if they recruit a bi-directional associative memory to associate sound patterns with articulatory gestures. Even more interest- ingly, the emergent vowel systems have the same statistical distribution as observed in human languages (see figure 1 from [12]). So this is a clear example where a recruitment approach has yielded surprisingly powerful explanations for a universal trend in human languages. They are more plausible than nativist explanations because crucial features of the vocal tract, such as a lowering of the larynx or the fine-grained control of the tongue, can just as well be explained through non-linguistic functional pressures, such as walking upright [1] or food manipulation before swallowing, and because detailed simulations of the Neanderthal vocal tract (with no lowered larynx and a more limited tongue flexibility) have shown that it is still possible to sustain a rich enough sound repertoire to build a language which is phonetically as complex as human languages today [4].

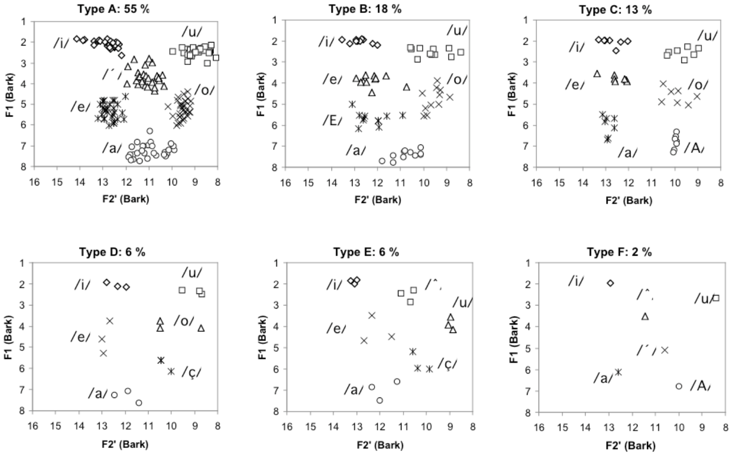

Fig. 1. Overlay of 100 different experiments in the self-organisation of vowels. The resulting vowel systems fall into different classes with similar characteristics as human languages and following a similar frequency distribution. Only the classes with 6 vowels are shown.

(2) Another source of support for the recruitment theory could come from evidence that similar brain areas are involved for dealing with linguistic and non-linguistic tasks, so that if there is a linguistic deficiency, there will be a related cognitive deficiency and vice-versa. This seems relatively easy. Whereas trying to find a gene that is unique for language or a brain impairment that exclusively affects language has proven to be like the search for a needle in a haystack, there exists overwhelming evidence of important correlations between linguistic and cognitive impairments. Brain imaging studies show that a verbal task always involves the activation of many other brain areas and that the areas supposedly genetically specialised for language are also active in non-linguistic tasks or that the functionalities of these areas can be taken over by others - possibly shifting from one hemisphere to another one. Even in the case of the FOXP2 gene, it is known that it is not uniquely relevant for language but plays an important role in the development of many areas of the brain and leads to many other impairments if affected, particularly in the motor domain. So the function of the gene could be much more generic, such as regulation of postmigratory neuronal differentiation [14].

(3) We can also construct mathematical arguments why certain strategies would be adopted by a population, as long as they yield better communication systems. This should be entirely feasable because before evolutionary game theory was adopted by biologists for studying genetic evolution, game theory was

already widely used by economists for studying how (rational) agents come to adopt the best strategy for achieving a specific task and the mathematical arguments used in evolutionary game theory can be carried over almost directly to multi-agent systems with peer-to-peer imitation and learning instead of genetic transmission [28]. Techniques from complex systems science, such as network analysis, are beginning to yield proofs why particular microscopic behaviors of agents lead to global coherence.

(4) Finally, we can engage in the same sort of computational simulations as have been carried out to examine the genetic evolution of language. In other words, we can examine how a stable communication system might emerge from repeated pairwise interactions of agents - but now without any genetic coding of the strategies for doing so, nor of the language systems themselves. We can then examine whether such systems have the major design features of human language, such as bi-directional use of signs, compositionality, marking of predicate-argument structure, explicit marking of perspective, tense-aspect-mood systems, etc. Work in my group has almost entirely focused on this kind of effort.

It is important to realise that these mathematical and computational models are a necessary complement to neurobiological or psychological evidence. Brain imagining or brain impairments can help us to see that a particular area with a known function has been recruited for language, for example, it may show that the parietal-temporal-occipital junction is active both during navigation and during the parsing of sentences with perspective reversal. But this only demonstrates that recruitment took place, it does not show why. This can only be done by comparing communication systems that have recruited this capacity to those that have not. It is only when a particular strategy leads to more communicative success, greater expressive power and/or greater cognitive economy that the strategy can be expected to survive.

In the next sections of this article I give three examples to illustrate more concretely how we can investigate the recruitment theory through computational simulations and robotic experiments. The structure of the experiments is always the same: We focus on a particular feature of language, identify mechanisms and strategies that may give rise to this feature in an emergent communication system, set up expeririments where agents endowed with these strategies play situated language games, and then test what difference the presence of this feature makes. The examples that follow examine quite different aspects of language: the establishment of linguistic conventions, the expression of predicate-argument structure, and the marking of perspective.

5 The Naming Challenge

Clearly every human language has a way to name individual objects or more generally categories to identify classes of objects. Computer simulations have already been carried out to determine what strategy for tackling this naming challenge could have become innate through natural selection [20] or how a shared lexicon could arise through a simulation of genetic evolution [7]. Although it is conceivable that an optimal strategy for acquiring a lexicon might have become genetically innate, it is highly unlikely that specific lexicons are genetically coded in the case of human languages, which have clearly language-specific lexicons of several tens of thousands of words which are continuously changing. The recruitment theory argues instead that each agent should autonomously discover strategies that allows him to successfully build up and negotiate a shared lexicon in peer-to-peer interaction and that the emerging lexicon is a temporal cultural consensus which is culturally transmitted. Various computer simulations (starting from [21] and [53]) have shown that the latter is entirely feasable and strategies have been found (by human designers) that are sufficiently robust to use them on situated em- bodied robot agents building up a common lexicon for perceptually grounded categories [40]. Research is also proceeding on how agents could discover such strategies themselves in an epigenetic recruitment process.

Here are some results of these experiments. They are based on the so called Naming Game [37] which is played by agents from a population taking turns being speaker and hearer. In each game, the speaker chooses a topic (usually from a subset of the possible objects which constitutes a context) and then looks up a word to name this topic. The hearer only gets the name and then has to guess which topic was intended. The game is a success if the hearer points to the topic that was originally chosen by the speaker. The game can fail for a variety of reasons and agents then repair their lexicons after a failure: The speaker can invent a new word if he has no way yet to name the topic, the hearer can adopt an unknown word because the speaker corrects a wrong choice in case of failure, and speaker and hearer can change their opinion about which word is most common for naming an object.

One mechanism that can be recruited for this game is a bi-directional memory which has weights between objects and their names, bounded between 0.0 and 1.0. This kind of memory is clearly very generic and can be used for storing all sorts of information [23]. Given this type of memory, many strategies are still possible:

1. If there are multiple choices (more than one name for the same object or more than one object for the same name), which one should be preferred?
2. Should a new word be invented each time the speaker fails to find a word?
3. Should the hearer always adopt an unknown word from the speaker, and if a new association is added to the lexicon, what should the initial weight be?
4. If the game succeeded, what should be the change in the association that was used by speaker and/or hearer?
5. If the game succeeds, should speaker or hearer do anything to the weights of competing associations, i.e. those with the same name but associated with another object, or conversely those with the same object but with another name?
6. If the game failed, what should be the change to the association that was used, if any?

Figure 2(top) shows what happens with a strategy where agents always choose the association with the highest weight, initialise new associations with

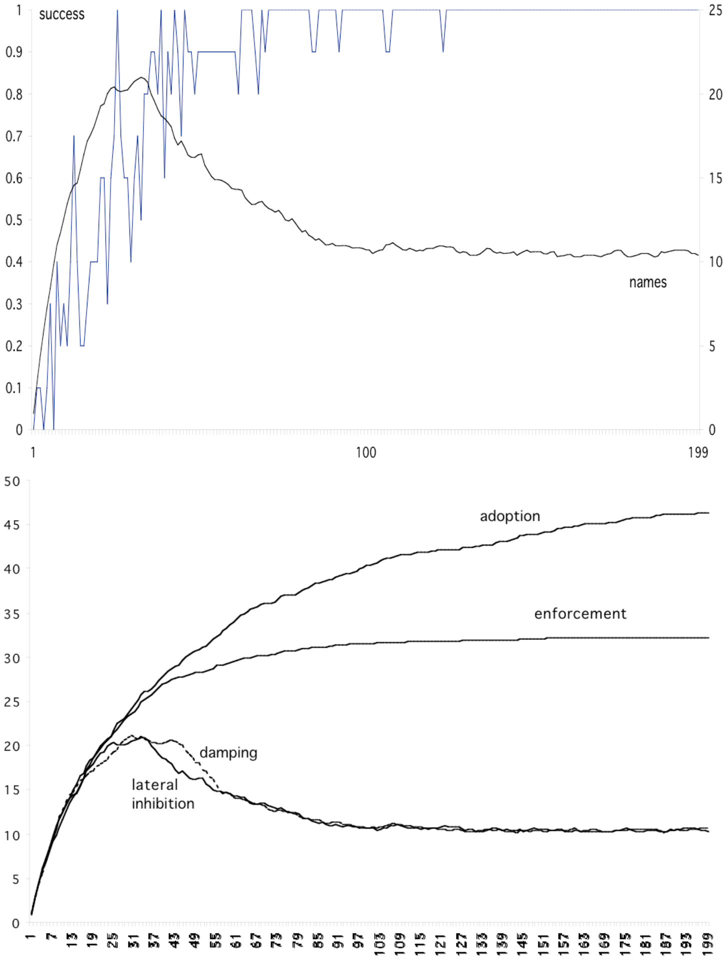

Fig. 2. Top: Evolution of communicative success and lexicon size over 2000 language games by a population of 10 agents naming 10 objects. Success quickly increases and the lexicon becomes optimal. Bottom: evolution in lexicon size for four different strategies. From top to bottom: adoption, enforcement, damping, lateral inhibition. Only the latter two strategies lead to an optimal lexicon.

0.5, use lateral inhibition in case of success, meaning that the weight of the used association is increased with 0.1 and the competitors decreased with 0.2, and decrease the used association in case of failure with 0.1. For a population of 10 agents who have to agree on names for 10 objects, this leads remarkably quickly to almost total success after about a 1000 interactions (that is roughly 200 per agent) and to an optimal lexicon with one name for every object. Figure 2(bottom) compares this strategy with a few others, focusing on the size of the lexicon, which impacts the time it takes for agents to reach consensus and the amount of memory they need to store a lexicon. Adoption means that there is no lateral inhibition and no decrease on failure. Enforcement means that there is an increase in case of success but no decrease of competitors and no decrease on failure. Lateral inhibition means that there is both enforcement and lateral inhibition but no decrease on failure. And finally damping means that there is not only enforcement and lateral inhibition but also a decrease on failure. All strategies lead to successful communication, but we see that only for the two last cases, inventory size is optimal. Many more experiments of this kind have been conducted, testing robustness against errors in signal transmission or feedback, flux in the population, increased homonymy, etc. [38].

It becomes even more interesting when agents do not have to name individual objects with proper names but perceptual categories for discriminating the topic from other objects in the context, so that there is no longer certainty and direct feedback about the meaning (as in Quine's famous Gavagai example). However simulations have shown that it is entirely possible to set up a semiotic dynamics where agents self-organise from scratch a shared vocabulary and a shared set of perceptual categories by adopting weights for both the categorical repertoire and the lexicon in the process [45]. It is particularly interesting that agents only arrive at a shared concept repertoire if concept development is coupled to language development (see figure 3 from [45]).

We learn from these experiments that (1) a variety of strategies for the naming challenge are effective but some are better than others, for example because they lead to a smaller lexicon, (2) when the agents apply these strategies they arrive at a shared lexicon through self-organisation and this lexicon is maintained even if there is an in- and outflux in the population, and (3) the key mechanism that had to be recruited is a bi-directional associative memory, even though there are still many possible strategies on how it should be used.

6 Expressing Predicate-Argument Structure

We now look at a second example: the marking of predicate-argument structure, i.e. who is doing what to whom. Marking predicate-argument structure is clearly a strong universal tendency in human languages, although many different ways have evolved to do so. Some languages (like Latin or Russian) use a complex system of cases and morphosyntax, others (like Chinese or Japanese) use a system of particles, still others (like English) mainly rely on verbal Subject-DirectObject-IndirectObject patterns with additional prepositions [3]. There

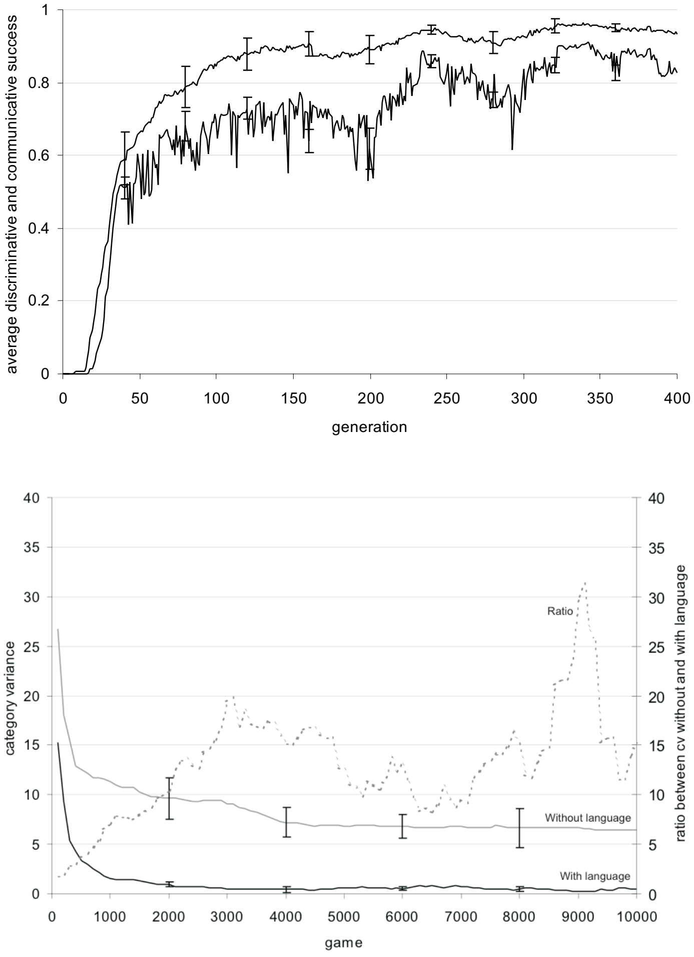

Fig. 3. The top figure shows discriminative (top curve) and communicative (bottom curve) success in a population building up concept repertoires and language conventions at the same time. Bottom: Ratio of category variance between the concept repertoires in a population of agents and category variance with (bottom) and without (top) language. When there is no coordination through language, the individual concept repertoires do not converge.

is moreover a general tendency that specific predicateargument relations are not expressed with ad hoc markers. Human languages have universally developed a more abstract intermediary layer in which a specific predicate-argument structure is first categorised in terms of more abstract frames, such as physical-transfer+agent+object+recipient, and these abstract roles are mapped onto a system of abstract markings (like nominative/ accusative/dative or subject/direct-object/indirect-object), so that there is a tremendous reduction in terms of the number of markers that are required [15], [18].

We have been carrying out various experiments to explore the expression of predicate-argument structure, using embodied agents which perceive the world through digital cameras. Agents are given access to real world scenes such as the ones shown in figure 4, in which puppets enact common events like the movement of people and objects, actions such as push or pull, give or take, etc. The recorded visual images are processed by each agent using a battery of machine vision algorithms that segment objects based on color and movement, track objects in real-time, and compute a stream of low-level features indicating which objects are touching, in which direction objects are moving, etc. These low-level features are input to an event-recognition system that uses an inventory of hierarchical event structures and matches them against the data streaming in from low-level vision, as described in detail in [44]. Using the world models resulting from these perceptual processes, agents then engage in description games, where the speaker describes one of the events in the most recent scene and the game is a success if the hearer agrees that this event indeed occurred. As before, the population starts without any pre-programmed language.

Fig. 4. Typical scene used in our case grammar experiments. The scenes are enacted with puppets and evoke typical interactions between humans and physical objects. A game succeeds if the hearer agrees that the event described by the speaker occurred in the most recent scene. In this case “Jill slides a block to Jack” is a possible description.

We have explored the hypothesis that agents mark predicate-argument structure because it reduces complexity in semantic interpretation [43]. Interpreting a meaning structure M_h with respect to a world model W_h with d possible objects is a function of the maximum number of possible assignments for a given meaning M_h with m variables, and hence it is equal to O(d^m). Searching through this set to find the assignment(s) that are compatible with Wh is therefore exponential in the number of variables and hence, reducing the number of variables by communicating which ones are equal is the most effective way to drastically reduce the complexity of semantic interpretation. This suggests the following strategy: Before the speaker renders a sentence, he may re-enter it into his own language system, parsing and interpreting the sentence to simulate how the hearer would process it (which is possible because every agent develops both the competence for being speaker and for being hearer). The speaker can thus compare what he originally wanted to express with what would be interpreted, and hence detect complexities in semantic interpretation that could be eliminated by introducing grammar. The hearer goes through a similar process: He first tries to interpret the sentence as well as possible, given his own perception of the scene, and then attempts to pick up which clues were added by the speaker to make the semantic interpretation process more efficient in the future. The precise implementation of the required mechanisms (particularly for grammatical processing) is too complex to describe in detail here (see [42]). But our experiments show convincingly that a shared system of case markers effectively arises when agents use this strategy (figure 5).

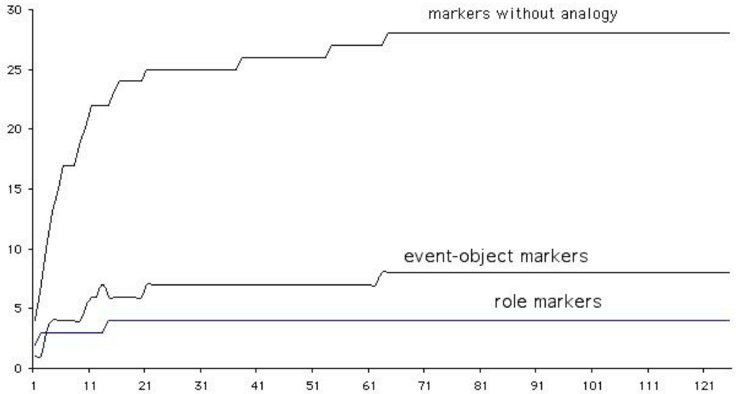

Fig. 5. The graph compares two strategies for marking predicate-argument structure in description games played by a population of agents about dynamical real world scenes: The top graph shows the number of specific case markers without recruiting analogy and the two bottom graphs show the specific case markers as well as the more generic role markers when analogy is recruited. The same series of scenes was considered in both cases. The recruitment of analogy clearly results in great economy with respect to the number of markers.

Next we endowed agents with an additional mechanism: The ability to establish an analogy between two events. This is a well-established generic cognitive component with wide applicability in a variety of tasks. When agents need to mark predicate-argument structure, they can now first try to find whether a convention to express an analogous event already exists. If so, the predicate-argument structures can be generalised as playing a common semantic role and the implicated words and markers can be categorised as belonging to the same parts of speech, thus giving rise to a new grammatical construction which can from now on be applied to other analogous events. As a side effect of using analogy, agents build up semantic and syntactic categories, similar to the way this is done in memory-based language learners [11]. As results in figure 5(right) show, this strategy drastically reduces the number of markers, which not only leads to a reduction of memory and processing time in rule matching, but also helps agents to reach a consensus more quickly and achieve communicative success for situations never encountered before.

We learn from these experiments that (1) the marking of predicate-argument structure can be explained as the outcome of strategies that attempt to reduce the computational complexity of semantic interpretation, (2) when agents apply these strategies, they arrive at shared grammatical systems, such as a grammar of case, in a process of cultural negotiation similar to the way lexicons have been shown to self-organise, (3) the recruitment of analogy for re-using existing structures automatically leads to an intermediary layer of abstract semantic and syntactic categories, which becomes progressively coordinated among the agents.

7 Perspective Marking

Another clear universal tendency of human speakers is to use not only their own perspective but also that of the hearer in order to conceptualise a scene, and to mark explicitly from what perspective the scene is viewed. For example, in English one can make a distinction between: your left and my left to mark whether the position is seen as left from the hearers perspective or from the speakers own perspective. German uses prepositions like herein and hinein depending on whether the required direction of movement is towards the speaker or away from the speaker, similar to English come and go. We now look at experiments with autonomously moving robots designed to explain why this universal tendency might occur [47].

The robot agents in these experiments are even more sophisticated than those used in the earlier predicate-argument experiments[1] They are capable of autonomous locomotion and vision-based obstacle avoidance, and maintain a real-time analog model of their immediate surroundings based on visual input (see figure 6 (left)). Using this vision-based world model, the robots are able to detect and track other robots as well as orange balls using standard image processing algorithms (see figure 6 (right)). Furthermore, the robots have been

---

[1] They are the Sony dog-like AIBO entertainment robots running programs designed specifically for these experiments.

endowed with mechanisms to segment the flow of data into distinct events and they have a short term memory in which they can store a number of past events. The robots play again description games, describing to each other the most recent event. The language game is a success if, according to the hearer, the description given by the speaker not only fits with the scene as perceived by him but is also distinctive with respect to previous scenes.

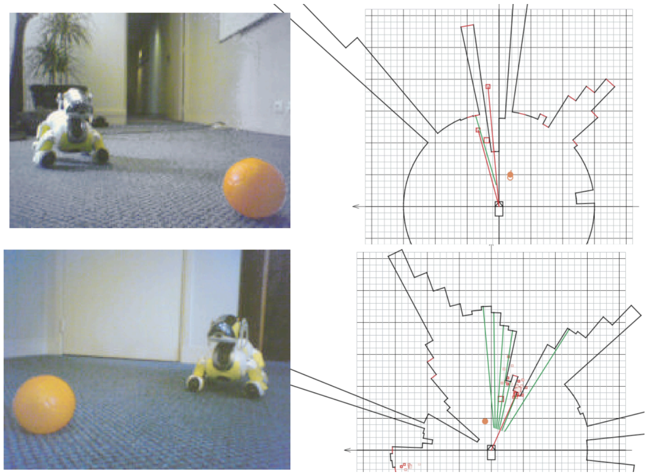

Fig. 6. The left images shows two robots (top and bottom) both perceiving the same orange ball but from different points of view. The right figures show the analog self-centered world models built up in real time by each robot. The self-position is located in the middle below and the self-orientation is North. The black lines indicate estimated positions of surrounding obstacles. Dots indicate estimated positions of the ball and the other robot.

At first the agents could recruit the same sort of strategies as before. They acquire categories in the form of discrimination trees (as in [46]), although other mechanisms for categorisation (e.g. Radial Basis Function networks or Nearest Neighbour Classification) would work equally well. They also use again a bi-directional associative memory with lateral inhibition and damping (as in the Naming Game experiment discussed in section 5). To support compositional coding (where more than one word can be used to cover the meaning) agents use a pattern decomposition system that combines partial matches from multi-word sentences [52].

Figure 8 shows the results when a population of ten agents applies these strategies in a series of 5000 one-on-one language games. We have first done an experiment where both agents use exactly the same perception of the world in the game. This allows us to establish whether the proposed mechanisms indeed work. Figure 8 (left) shows that this is indeed the case. Commu- nicative success rises to hover around 90%, despite the fact that none of the agents had any categories or

Fig. 7. The left images illustrated the world models computed by the speaker (top) and the hearer (bottom). The self-position (the rectangle below in the middle), the estimated position of the other robot (the other rectangle), the balls begin and end position, as well as the ball trajectory are shown. The right images show the world models of speaker (top) and hearer (bottom) after egocentric perspective transformation, so that they now illustrate how one robot hypothesises the perception of the world by the other.

words to start with. Residual communicative error is due to perceptual problems (for example one robot does not even see where the other robot is) and not to incompatibilities of linguistic inventories or ways of conceptualising the world. Figure 8 (left) shows that the size of the lexicon stabilises to about 50 words, reached when they are sufficient to cope with the communicative challenges in this particular setting.

We next do an experiment where agents are truely embodied and hence have different perceptual views on the same scene. Figure 8 (right) shows that there is still some success due to situations where the viewpoint both agents have on the scene is sufficiently similar. However the communication system does not really come off the ground. Success is around 50% and the lexicon is almost three times as large. This clearly shows that if situated embodied agents want to be successful in describing the movement of objects vin the space around them, they will have to recruit additional mechanisms.

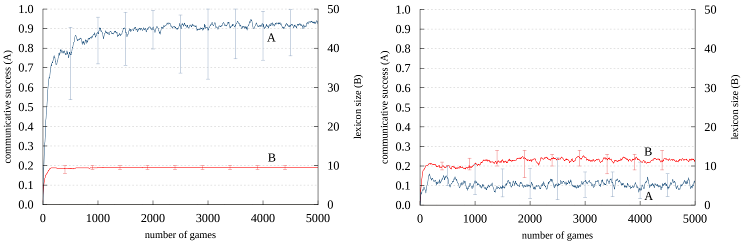

Fig. 8. The graphs shows communicative success (upper curve blue) and the size of lexical inventories (lower curve red) over 5 runs for 5000 language games each. The left graphs show the case where agents share the same perception of the world and the right graphs where they have each a different viewpoint. We see a dramatic drop in communicative success. The lexicon is no longer converging in the latter case.

We now endow agents with the capacity of egocentric perspective transformation, which is a well established quasi-universal mechanisms in human cognition [17]. Egocentric perspective transformation allows agents to reconstruct a scene from the viewpoint of another agent. It requires that they first detect where the other agent is located (according to their own perception of the world) and then perform a geometric transformation of their own world model (see the example in figure 7 (right)). Inevitably, an agents reconstruction of how another agent sees the world will never be completely accurate, and may even be grossly incorrect due to unavoidable misperceptions both of the other robots position and of the real world itself. The sensory values obtained by the robots should not be interpreted as exact measures (which would be impossible on physical robots using real world perception) but at best as reasonable estimates. This type of inaccuracies is precisely what a viable communication system must be able to cope with and robotic models are therefore a very good way to seriously test and compare strategies and the mechanisms that implement them.

Figure 9 (left) shows that agents clearly increase communicative success when they recruit egocentric perspective transformations for conceptualisation. Even without explicitly marking perspective, success dramatically increases because agents try out what makes sense, first from their own perspective and then from the others perspective. Here as in all our other experiments, agents use an inferential coding strategy which means that part of the meaning can be inferred or gleaned from the shared situation [36]. The lexicon is almost the same size as in the first experiment and communicative success is steadily around 80%. In the next experimental run figure 9 (right), agents not only use egocentric perspective transformation for conceptualisation but also mark perspective when necessary, i.e. the adopted perspective (self or other) is part of the meanings they convey. We see that communicative success is similar. The size of the lexicon is bigger but beginning to settle to a more optimal one thanks to the compositional coding of perspective.

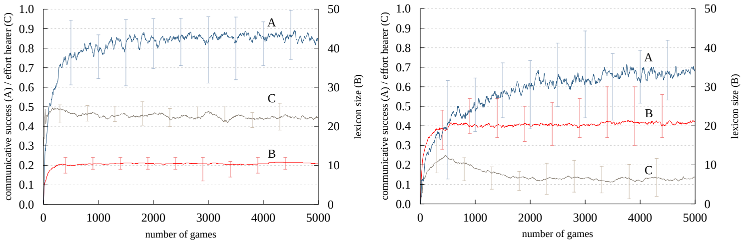

Fig. 9. Communicative success and lexical inventory size for different world perception with egocentric perspective transformation without explicit marking (left), and with egocentric perspective transformation and the marking of perspective (right). The right graph shows not only communicative success and inventory size but also the decreasing effort involved in perspective transformation.

Figure 9 (right) also shows the gain in cognitive economy. The bottom curve shows the amount of effort involved in computing different perspectives. In the experiment where no perspective is marked, agents always need to take the other perspective to check which perspective the speaker might have used. When perspective is marked, agents can optimise this computation and only do it when perspective reversal is explicitly signalled. We therefore see a decrease in effort (and hence increase in cognitive economy). We conclude that embodied and situated agents, which have different viewpoint on a scene, can reach significantly more success in communication when they recruit egocentric perspective transformation for conceptualising the scene and even more success if they also mark explicitly the adopted perspective in their language so that they know whether to apply perspective transformation or not. This justifies the adoption of this strategy and hence explains why it is universally found in human languages.

8 Discussion

The present article discussed the recruitment theory of language origins. This theory argues that human brains are capable to dynamically recruit various cognitive mechanisms and configure them into strategies for handling the challenges of communication in particular environments. The configurations are retained if they increase communicative success and expressive power while minimising the effort involved (processing time, memory resources, etc.). Each strategy has an impact on the emergent language and hence it gets culturally fixated. We discussed a number of very concrete experiments with computer simulations and robotic models that attempt to substantiate the recruitment theory. Each experiment focuses on a specific universal feature of language and tries to show that it is the consequence of the recruitment of a particular strategy and the cognitive mechanisms that can implement this strategy. If the adoption of the strategy leads to a better communication system, then we have a functional explanation why a human language may have this particular feature.

The experiments reported here all compare different strategies and mechanisms. They show that communication systems with similar features as in human natural languages can emerge from the collective behaviors of individuals and that the adoption of some strategies leads to better communication systems compared to others. But these experiments do not attack yet the problem how the recruitment process itself might work in individuals, nor how individual recruitment processes are influenced by solutions already adopted by others. Computational simulations and robotic experiments on this issue are progressing and they rest on earlier work in genetic programming [24], classifier systems [19], and other forms of adaptive machine learning in multiagent systems. Research can rely partly on a strong tradition in ethology to study behavioral decision-making in animals, which often appears to reach remarkably optimal performance [29].

As mentioned earlier, researchers in favor of innate linguistic structure split between adaptationists (such as Pinker, a.o.) who claim that it is the (communicative) function of language which acts as the selectionist drive because it directly influences fitness and hence the spreading of genes, and non-adaptationists (such as Chomsky, Gould, a.o.) who claim that language does not appear to be designed well or primarily for communication and hence that other factors must have played a role to genetically shape the language organ. Similarly there are researchers who are generally in favor of the language-as-a-complex-adaptive-system view but do not take communicative function to be the main force that guides the process of (cultural) selection. One example is the work of Kirby and collaborators [5]. Instead of communicative function, they take the challenge of cultural transmission (how a language system can be learned by the next generation) as the main driver for the emergence of linguistic structure and they have obtained a number of significant results using their iterated learning framework, including the selection of compositional versus holistic coding. The recruitment theory argues that communicative function, including the sensori-motor embodiment and the real world environments and ecologies constraining the behavior of the population, shape language. Learning still plays a significant role in the sense that conventions which are not learnable cannot be picked up by the hearer and hence will not propagate in the rest of the population, but learning is incorporated as part of peer-to-peer interactions instead of transmission from one generation to the next.

The recruitment theory is at this point still a research program instead of a fully complete and established theory. Many features of language remain to be investigated and much more research is needed on cognitive mechanisms. The experiments briefly reported here are very difficult to do, requiring a team working over several years. However, they yield very clear and solid results, thus establishing a growing body of functional explanations for human language. Much more research is needed on the neurological basis for the recruitment theory. It requires taking the opposite view from what is now common in brain science. Rather looking to see what is unique for language (genes or neural subsystems) we need to find out what is common between linguistic and non-linguistic tasks.

9 Acknowledgement

The research reported here is currently funded by the Sony Computer Science Laboratory in Paris with additional funding from the EU FET-ECAgents project 1170. Many researchers have participated in the experiments reported here, most of them as part of their Ph.D. thesis projects at the Artificial Intelligence Laboratory of the University of Brussels (VUB): Paul Vogt and Edwin de Jong participated in the early discoveries of the naming game, Frederic Kaplan, Angus McIntyre and Joris Van Looveren were key collaborators for the early Talking Heads experiments, which continued with a focus on the co-development of conceptual and linguistic repertoires by Tony Belpaeme and Joris Bleys. Pierre-Yves Oudeyer and Bart de Boer have successfully explored the emergence of sound repertoires through imitation games. Nicolas Neubauer and Joachim De Beule collaborated on the development of the Fluid Construction Grammar formalism and the predicate-argument experiments. Martin Loetsch and Benjamin Bergen played key roles in the perspective reversal experiments.

References

1. Aiello, L. C. (1996) Terrestriality, bipedalism and the origin of language. Runciman, W.G. , J. Maynard-Smith, and R.I.M. Dunbar (eds.) Evolution of Social Behaviour

Patterns in Primates and Man, Oxford: Oxford University Press. p. 269-89.

2. Bickerton, D. (1984) The language bioprogram hypothesis. Behavioral and Brain Sciences 7, 17–388.
3. Blake, B.J. (1994) Case. Cambridge University Press, Cambridge UK.
4. Boe, L-J., Heim, J.-L., Honda, K. and Maeda, S. (2002) The potential Neandertal vowel space was as large as that of modern humans. Journal of Phonetics, 30(3): 46584.
5. Brighton, H. and S. Kirby (2001) The survival of the smallest: stability conditions for the cultural evolution of compositional language. In J. Kelemen and P. Sosik (Eds.) Advances in Artificial Life. pp592601 Berlin: Springer-Verlag
6. Briscoe, T. (ed.) (2002) Linguistic Evolution through Language Acquisition: Formal and Computational Models. Cambridge University Press, Cambridge, UK.
7. Cangelosi A., D. Parisi (1998). The emergence of a language in an evolving population of neural networks. Connection Science, 10(2), 83-9
8. Cangelosi, A. and D. Parisi (eds.) (2001) Simulating the Evolution of Language. Springer-Verlag, Berlin.
9. Chomsky, N. (1988). Language and Problems of Knowledge. Cambridge, MA: MIT Press.
10. Croft, W. (2001) Radical Construction Grammar: syntactic theory in typological perspective. Oxford University Press, Oxford.
11. Daelemans W. (2002) A comparison of analogical modeling to memorybased language processing.- In: Skousen R. (ed) Analogical modeling. Amsterdam, Benjamins, 2002, p. 157-179
12. De Boer, B. (2000) Self organization in vowel systems. Journal of Phonetics 28(4): 441465.
13. Enard et al (2002) Molecular evolution of FOXP2, a gene involved in speech and language, Nature 418, 869 - 872.
14. Ferland, R.J., T. J. Cherry, P. O. Preware, E. E. Morrisey, and C. A. Walsh (2003) Characterization of Foxp2 and Foxp1 mRNA and Protein in the Developing and Mature Brain. The Journal of Comparative Neurology 460:266279 (2003)
15. Fillmore, Charles J. (1976): Frame semantics and the nature of language; in Annals of the New York Academy of Sciences: Conference on the Origin and Development of Language and Speech, Volume 280 (pp. 20-32)
16. Tecumseh Fitch, W., M. D. Hauser and N. Chomsky (2006) The Evolution of the Language Faculty: Clarifications and Implications. Cognition (in press).
17. Flavell, J.H., 1992. Perspectives of perspective taking. In: Beilin, H., Pufall, P. (Eds.), Piagets Theory: Prospects and Possibilities. Erlbaum, Hillsdale, N.J.
18. Goldberg, A.E. (1995) Constructions.: A Construction Grammar Approach to Argument Structure. University of Chicago Press, Chicago
19. Holland, J.H. (1992) Adaptation in Natural and Artificial Systems. MIT Press, Cambridge, Massachusetts.
20. Hurford, J. (1989) Biological evolution of the Saussurean sign as a component of the language acquisition device. Lingua 77, 187-222.
21. Hutchins, E. and B. Hazlehurst (1995) How to invent a lexicon: The development of shared symbols in interaction. In N. Gilbert and R. Conte (eds.) Artificial societies: The computer simulation of social life (p. 157- 189). UCL Press, London.
22. Iachini, T. and R. Logie (2003) The Role of Perspective in Locating Position in a Real-World, Unfamiliar Environment, Appl. Cognit. Psychol. 17:904.
23. Kosko, B. (1988) Bidirectional Associative Memories. IEEE Transactions on Systems, Man, and Cybernetics, 18, pp. 49-60, 1988

24. Koza, J. (1992) Genetic Programming: On the Programming of Computers by Means of Natural Selection. The MIT Press, Cambridge Ma.
25. Kuhl, P. K., Meltzoff, A. N. (1997). Evolution, nativism, and learning in the development of language and speech. In M. Gopnik (Ed.), The inheritance and innateness of grammars (pp. 7-44). New York: Oxford University Press.
26. Lieberman, P. (1975) On the Origins of Language. Macmillan, New York.
27. Lindblom, B., P. MacNeilage, and M. Studdert-Kennedy (1984) Selforganizing processes and the explanation of phonological universals. In: B. Butterworth, B. Comrie and O. Dahl (eds.), Explanations for Language Universals, 181203. Berlin: Walter de Gruyter.
28. Matsen, F.A., and M.A. Nowak (2004) Win-stay, lose-shift in language learning from peers. PNAS, Vol 101(52) 18053-18-57.
29. McFarland, D.J. (1977) Decision making in animals. Nature 269: 15-21.
30. Minett, J. W. and Wang, W. S-Y. (2005) Language Acquisition, Change and Emergence: Essays in Evolutionary Linguistics. City University of Hong Kong Press: Hong Kong.
31. Nowak, M.A., N. Komarova, P. Niyogi (2001) Evolution of universal grammar. Science, 291:114-118, 2001
32. Pinker, S. (2004) Language as an Adaptation to the Cognitive Niche. In: Christiansen, M.H. and S. Kirby (eds) Language Evolution: The States of the Art. Oxford University Press, Oxford. Chapter 2.
33. Pinker, S. and R. Jackendoff (2005) The Faculty of Language: Whats Special about it? Cognition, in press.
34. Oudeyer, P-Y. (2005) The Self-Organization of Speech Sounds, Journal of Theoretical Biology, Volume 233, Issue 3, pp.435449
35. Schwartz, J. L., Boe, L. J., Vallee N. and C. Abry (1997) Major trends in vowel system inventories. Journal of Phonetics 25, pp. 233253
36. Sperber, D. and Wilson, D. (1995) Relevance communication and cognition. Blackwell Pub., London.
37. Steels, L. (1995) A Self-organizing Spatial Vocabulary. Artificial Life 2(3), pp. 319-332.
38. Steels, L. and F. Kaplan (1998) Stochasticity as a source of innovation in language games. In C. Adami and R. Belew and H. Kitano and C. Taylor, editors, Artificial Life VI. Los Angeles: MIT Press.
39. Steels, L. (2000) Language as a complex adaptive system. Lecture Notes in Computer Science. Parallel Problem Solving from Nature - PPSN-VI. Volume Editor(s) : Schoenauer et al., Springer-Verlag, Berlin.
40. Steels, L. (2003) Evolving grounded communication for robots. Trends in Cognitive Science. Volume 7, Issue 7, July 2003 , pp. 308-312.
41. Steels, L. (2003) Intelligence with representation. Philosophical Transactions Royal Society: Mathematical, Physical and Engineering Sciences, 361(1811):23812395
42. Steels, L. (2004) Constructivist Development of Grounded Construction Grammars Scott, D., Daelemans, W. and Walker M. (eds) (2004) Proceedings Annual Meeting Association for Computational Linguistic Conference. Barcelona. p. 9-19.
43. Steels, L. (2005) What triggers the emergence of grammar? Dautenhahn, K. (ed.) (2005) Proceedings of the AISB 2005 conference. EELC satelite workshop. University of Herfortshire, Hatfield, UK.
44. Steels, L. and J-C. Baillie (2003) Shared Grounding of Event Descriptions by Autonomous Robots. Journal of Robotics and Autonomous Systems 43, 2003, 163-173.
45. Steels, L. and T. Belpaeme (2005) Coordinating Perceptually Grounded Categories through Language. A Case Study for Colour. Behavioral and Brain Sciences. 24(6).

46. Steels, L., F. Kaplan, A McIntyre and J. Van Looveren (2002) Crucial factors in the origins of word-meaning. In Wray, A., editor, The Transition to Language, Oxford University Press. Oxford, UK, 2002.
47. Steels, L., M. Loetzsch and B. Bergen (2005) Explaining Language Universals: A Case Study on Perspective Marking. [submitted]
48. Stromswold, K. (2001) The Heritability of Language: a Review and Metaanalysis of Twin, Adoption, and Linkage Studies. Language, 77(4), p. 647-723.
49. Szathmary, E. (2001) Origin of the human language faculty: the language amoeba hypothesis. In: J. Trabant and S. Ward (eds) New Essays on the Origins of Language. Trends in Linguistics, Studies and Monographs Vol.133. p. 41-55.
50. Tomasello, M. (1999). The cultural origins of human cognition. Harvard University Press, Cambridge, MA.
51. Thornton, R. and K. Wexler (1999) Principle B, VP Ellipsis, and Interpretation in Child Grammar. The MIT Press, Cambridge Ma.
52. Van Looveren, J. (1999) Multiple word naming games. In Proceedings of the 11th Belgium-Netherlands Conference on Artificial Intelligence. Universiteit Maastricht, Maastricht, the Netherlands, 1999.
53. Werner, G. and M. Dyer (1991) Evolution of Communication in Artificial Organisms, in Langton, C. et al. (ed.) Artificial Life II. Addison- Wesley Pub. Co. Redwood City, p. 659-687
54. Wray, A., ed. (2002) The Transition to Language, Oxford University Press. Oxford, UK, 2002.
55. Zacks, J., B. Rypma, J.D.E. Gabrieli, B. Tversky, G. H. Glover (1999) Imagined transformations of bodies: an fMRI investigation. Neuropsychologia 37 (1999) 1029-1040.