ARTICLE IN PRESS

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Language Sciences xxx (2007) xxx–xxx

Language Sciences

www.elsevier.com/locate/langsci

Adaptive agent modeling of distributed language: investigations on the effects of cultural variation and internal action representations

Angelo Cangelosi *

Adaptive Behaviour and Cognition Research Group, School of Computing, Communications and Electronics, University of Plymouth, Drake Circus, Plymouth PL4 8AA, UK

Abstract

In this paper we present the ‘‘grounded adaptive agent’’ computational framework for studying the emergence of communication and language. This modeling framework is based on simulations of population of cognitive agents that evolve linguistic capabilities by interacting with their social and physical environment (internal and external symbol grounding). These models provide an integrative vision of language where the linguistic abilities of cognitive agents strictly depend on other social, sensorimotor, neural and cognitive capabilities. Here language is not seen as an isolated and dedicated symbol processing system, but rather as a heterogeneous set of artifacts implicated in cultural and cognitive activities. The proposed modeling approach is also closely related to embodied cognition theories of the grounding of language in the organism's perceptual and motor systems.

Keywords: Symbol grounding; Language evolution; Computational modeling; Neural networks; Embodied cognition

1. Introduction

Language has been typically seen as an internal processing system dedicated to the mapping of semantic representations into a set of words and the rules used to combine these into

* Tel.: +44 1752 232559; fax: +44 1752 232540.
E-mail address: acangelosi@plymouth.ac.uk

0388-0001/$ - see front matter
doi:10.1016/j.langsci.2006.12.026

sentences. Although language researchers acknowledge the social role of language as a communication system, their main focus has been on this hypothesized internal, specialized processing device. This view is typical, above all, of formal approaches to language analysis. Thus in generative linguistics, for example, language is viewed as an isolated and autonomous capability (e.g. Chomsky, 1975; Fodor, 1975; Burgess and Lund, 1997).

An alternative approach to language has been recently proposed. Here language is not seen as an isolated and dedicated symbol processing system, but rather as a heterogeneous set of artifacts implicated in cultural and cognitive activities (Cowley, 2005; Linell, this issue). Although not excluding the important internal processing capability, language is viewed as distributed. Attention is thus also given to how bodies and artifacts impact on cognitive and linguistic dynamics.

In this paper we propose a computational modeling approach to signaling and language that integrates the internal and external roles of language. This is the framework of grounded adaptive agent modeling (Cangelosi, 2004, 2005). It is based on simulations of population of cognitive agents that evolve linguistic capabilities by interacting with their social and physical environment. These models provide an integrative vision of language where the linguistic abilities of cognitive agents strictly depend on other social, sensorimotor, neural and cognitive capabilities. All these factors contribute to the emergence and establishment of communication and language. This modeling approach is closely related to psychologically plausible theories of the grounding of language in the organism's perceptual and motor systems including, for example, Perceptual Symbol Systems (Barsalou, 1999) and the action-grounding theory of language processing (Glenberg and Kaschak, 2002).

The integration of internal and external factors in language studies implies that language is grounded both into the agent's own cognitive representations (internal symbol grounding) *and* into the social and physical environment in which linguistic capabilities develop (external symbol grounding). These will be discussed below, together with reference to existing approaches to both kinds of grounding.

1.1. Internal symbol grounding

One of the fundamental issues in the design of computational cognitive models is known as the symbol grounding problem (Harnad, 1990; Cangelosi, 2005). This states that a psychologically plausible model of cognition and language must be based on a set of symbols (e.g. words) that can be intrinsically grounded in the agent's cognitive representation of the world. For example, a cognitive or agent model that can communicate with humans (or other agents) about plant and animal taxonomy must be able to autonomously link the words used to identify plants/animals to internal representations of plant and animal categories. These representations can be acquired during interaction with the entities in, for example, learning categories that "sort out" our external environment. The mechanism of internal symbol grounding is therefore based on the following chain of entities and representation:

$$\text{external entities} \Longleftrightarrow \text{internal representations} \Longleftrightarrow \text{symbols}.$$

These bi-directional links permit the external entities to influence internal representation and symbols while, simultaneously, symbols affect the way we represent our external world (Roy, 2005).

Please cite this article in press as: Cangelosi, A., Adaptive agent modeling of distributed ..., Lang. Sci. (2007), doi:10.1016/j.langsci.2006.12.026

Harnad suggests that the internal representations that mediate grounding are essentially categorical. That is, different members of the category of *dogs* share a common internal representation that is reliably associated with the word "dog". One important characteristic of these representations is that of categorical perception (Harnad, 1987). This refers to a process of warping the similarity space of internal categorical representations. When this happens, the perceptual (iconic) representations of a category are transformed to ensure that internal within-category differences diminish (within-category compression). Differences between iconic representations of the members of different categories are enhanced to make them appear more dissimilar (between-category expansion). Such phenomena of categorical perception are widespread in natural cognitive systems, and have been shown to occur in both animals (e.g. Zentall et al., 1986) and humans (e.g. Goldstone, 1994).

There are several models of how categorical representation can lead to the grounding of language. For many, hybrid symbolic-connectionist models are often seen as ideal candidates for solving the symbol grounding problem (Harnad, 1993; Sun, 2002). This is because a connectionist (neural network) component is able to perform the task of grounding basic symbols into perceptual and categorical representations. Input units receive iconic representation of the objects and, using hidden units, these are transformed to warped, categorical perception representations. A symbolic module, such as a rule-based system, would use categorical representations to perform symbol manipulation operations and high-level cognitive tasks (Miikkulainen, 1994). Instead of invoking hybrid systems based on connectionist architectures, Cangelosi (2005; also Cangelosi et al., 2000) focuses on neural networks that are able to (i) acquire categorical perception representations during category learning tasks, (ii) use these to ground the meaning of discrete symbols, and (iii) perform symbol manipulation tasks. Finally, Vogt (2002) proposes approaching the symbol grounding problem by using embodied agent models and cognitive robotics (Vogt, 2002). These methodologies permit the grounding of language in sensorimotor representations. While some robotic approaches incorporate use of connectionist networks, others use different control architectures. Such a robotic/agent approach not only deals with the internal grounding of language, but, as described below, also with the external grounding of symbols in social learning and interaction.

1.2. External symbol grounding

The external symbol grounding problem refers to the social and cultural grounding of language and communication (Cangelosi, 2006). This involves a wider interpretation of communication to include artifacts and cultural symbols and is thus consistent with a distributed view of language. The most important issue in the external symbol grounding problem is the role of social mediation in acquiring shared symbols that can be used for cultural and communicative purposes.

Embodied agent and robotics approaches to the grounding of language are ideal candidates for tackling the external symbol grounding problem (Steels, 1999, 2002; Cangelosi, 2001). These models always focus on a group, or population, of cognitive agents that form a shared set of communication symbols through direct negotiation. The language that develops lends itself to population thinking, because it affects a group of social agents. Agents have to perform a collaborative task, often based on a "language game" between a hearer and a speaker (Steels, 1999). This consists in the speaker's selecting a topic of communication through shared attention, producing an utterance to communicate its internal representation of the topic, and the hearer's comprehension of the utterance which leads to the word being incorporated in its own lexicon. The success/failure of the language game results in the update of the agents' lexicons, including, when necessary, the invention of new words. Some of these robotic models also involve human subjects who negotiate and contribute to the acquisition of shared lexicons (e.g. Steels and Kaplan, 1999; Hsiao et al., 2003).

External symbol grounding in robotics has tended to focus on the dynamics of social and cultural grounding. In short, there has been a tendency to see this as extending internal symbol grounding to the social community. As such, the focus has fallen on how words used in communication can lead to the emergence of shared lexicons. The distributed view would require the extension of the repertoire of ''symbols'' to other codes and cultural activities. Although models have focused on the emergence of artifacts (Ugolini and Parisi, 1999), little attention has been given to making their communicative role explicit.

2. Grounded adaptive agent models of distributed languages

Grounded adaptive agent models provide an integrative vision of language where the linguistic abilities of agents are grounded in both their own (internal) cognitive capabilities and (external) social and cultural mechanisms. Such an approach is compatible with a variety of modeling methodologies. Some use simulated robots interacting with (simulated) physical entities, while others use multi-agent simulations. In robotic models, communication results from dynamics that link the robot's physical body, its cognitive system and the physical and social environment. For example, some have focused on how robotic arms can be used in grounding sensorimotor processes (see, Marocco et al., 2003; Cangelosi and Riga, 2006). Others have used multi-agent models to simulate foraging agents that learn to communicate about foods (Cangelosi, 2001; Cangelosi and Harnad, 2000) and, indeed, yet others have used agents that perform object manipulation tasks (Cangelosi and Parisi, 2004; Hazlehurst and Hutchins, 1998). These multi-agent simulations model the agent and its environment with a degree of detail that is sufficient to contribute to the construction of emergent meanings.

Both robotic and simulation models focus on the emergence of simple communication systems and the evolutionary transitions from animal-like communication to combinatorial languages. In this paper we will review models of how multi-agent systems ground language. In Section 2.1 we will analyze one of these models to show how a population of grounded adaptive agents develop a shared compositional language. The lexicon exploits a combination of evolutionary mechanisms (e.g. the Baldwin effect) and cultural processes (e.g. parent–child language learning and the rate of cultural variation). In Section 2.2 we will describe analysis of the internal structure of the lexicon and underlying sensorimotor representations in a model of the emergence of proto-syntactic languages. Due to space limitations, we will only provide brief description of the models and results. Full details on these two simulations are available in Munroe and Cangelosi (2003), for the first model, and Cangelosi and Parisi (2004; Cangelosi, 2004) for the second.

An important assumption in discussing these computational models of the evolution of ''language'' is that the term ''language'' is used in the sense of *shared signaling system*. For example, in simulating the sensorimotor basis of ''verbs'' and ''nouns'', the model suggests that what an observer calls a ''verb''' is not an abstract entity associated with a lexical item. Rather, it is a signal that contributes to behavior because of how it is integrated with

Please cite this article in press as: Cangelosi, A., Adaptive agent modeling of distributed ..., Lang. Sci. (2007), doi:10.1016/j.langsci.2006.12.026

an agent's past history. In a verb-using (or verb recognizing) agents, the 'verb' is integral to what the agent does and/or perceives.

2.1. The evolutionary emergence of shared compositional languages

The model considers the case in which a population of autonomous agents live in a simple world and have to forage to be able to survive and reproduce. They are also endowed with a capacity (instinct) to produce sounds (signals) and to listen to and imitate vocalizations produced by other agents. The simulation model can help us answer some of the following questions: What kind of signaling might evolve? Will the agents be able to converge towards the use of a shared set of symbols (conventions)? What are the properties of the evolved signaling system, (e.g. in terms of compositional structure and linguistic categories)? How do phylogenetic mechanisms and ontogenetic learning interact to support the evolution of a shared signaling system?

2.2. The model and simulation setup

The model uses a mushroom world metaphor where simple agents forage and communicate while seeking food (Cangelosi and Harnad, 2000). An agent gains from identifying foods by using the visual features of mushrooms. In the environment there are three types of edible mushrooms and three poisonous toadstools. All mushrooms can be approached and eaten, and all toadstools are to be avoided. Each sub-type of edible mushroom can only be eaten properly if a specific action is chosen (e.g. wash/cut/squash). The agents' fitness is measured as the amount of energy gained at the end of a 'life'. They gain 1 energy point for each food they have eaten and lose 1 point for any poisonous items consumed.

An agent's behavior is controlled by a feed-forward neural network (multi-layer perceptron). The network contains three sets of input units: (i) three nodes to encode the position of the food, (ii) eighteen to encode the perceptual properties of the food, (iii) eight linguistic input units. The hidden layer has five nodes. The output layer will consist of two groups of units: (i) three encode the agents' actions (two to move in the environment and one to decide the wash/cut/squash action), and (ii) eight encode the linguistic output. The linguistic output nodes have two winner-takes-all clusters, one with two nodes, and another with six. Within each cluster, the node with the highest activation is set to 1 and the rest are set to 0. Each unit of a cluster encodes the activation of a "word". At any time step, two words are produced.

In the first stage of the simulation, a population of 80 agents is evolved according to their ability to collect mushrooms, using a genetic algorithm (Holland, 1975). The weights of the neural networks are encoded in the genotype. At each generation, the 20 agents with the highest fitness are selected. They reproduce by producing four offspring, each of whom is subject to random mutations in its 'genotype'. This first evolutionary stage evolution lasts 300 generations. In the next 100 generations (stage 2) agents communicate by receiving linguistic instructions from their parents. They must continue to forage successfully but, in this stage, gain access to visual features of the food only 10% of the time. Instead, they rely on their parent's linguistic descriptions (each parent only speaks to its own offspring). During each cycle of parent–child interaction, three different tasks are executed (Fig. 1). In the first (Listening), the parent sends out two signals to give a verbal descrip-

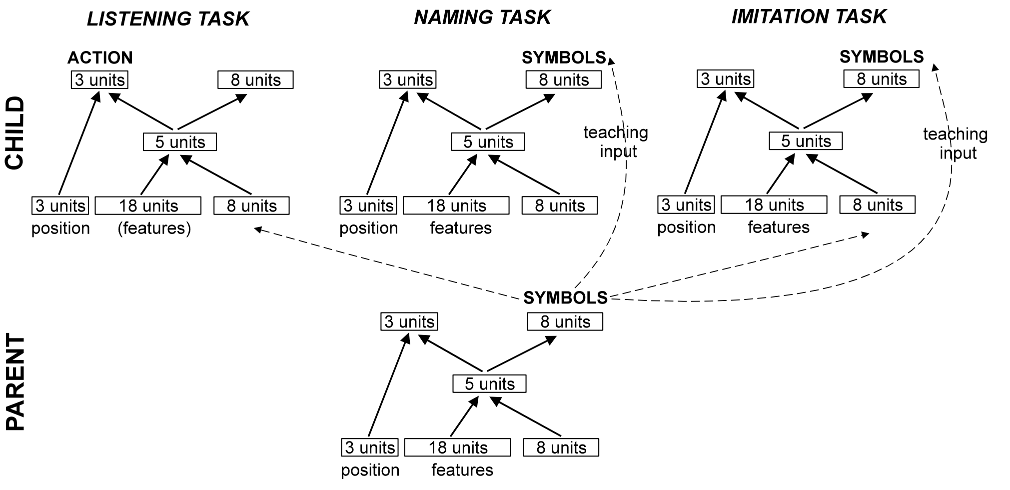

Fig. 1. Interaction and learning between parent (speaker) and child agent (listener). From Cangelosi (2001).

tion of food close to the child. When no visual information is provided, as in 90% of the cases, the child uses this linguistic information to decide what to do with the mushroom (approach/avoid and wash/cut/squash). In the second task (Naming), the listener also performs a naming task while looking at the mushroom's visual features and activating two output linguistic clusters. Through a backpropagation learning cycle (Rumelhart et al., 1986), the child corrects output using teaching input signals provided by the parent. In the third task (Imitation), the child performs a linguistic imitation task to learn, via back-propagation, to repeat parental description. This imitation leaning cycle is important to allow the hidden units to specialize for both the listening and the naming task (see Cangelosi et al., 2000).

2.3. Results on the emergence of compositionality

The structure of the lexicon in the simulations that successfully evolved linguistic communication at generation 400 can be classified into three types: (i) single-word languages; (ii) combinatorial languages; (iii) compositional languages. Out of the 7 (out of 8) populations that successfully use language to forage, one evolved a single-word lexicon, i.e. through the use of only the 6-unit cluster to differentiate between mushrooms (the 2-unit cluster constantly produces the same signal, therefore making it information-less). While two populations evolved combinatorial two-word pairings, their systems lacked compositionality because the signals were not systematically associated with identifiable meaning units (e.g. no signal was consistently associated either with "avoid" or "approach" or mushroom types). The final four populations developed a compositional system that resembled verb–noun languages.

An example of evolved compositional language is shown in Table 1. This reports the frequency distribution of signals selected by each parent to name all the mushrooms. In the 2-unit cluster, agents use each of the two nodes to indicate the poisonous and edible items respectively. These can be considered as proto-linguistic verbs that indicate the

Table 1
Frequency distribution of the use of communication signals in one population that evolved a shared compositional language

	6-Unit cluster						2-Unit cluster	
	A	B	C	D	E	F	*X*	*Y*
Edible 1						20		20
Edible 2	17		1		2			20
Edible 3				20				20
Toadstool 1			16			4	20	
Toadstool 2			20				20	
Toadstool 3			20				20	

The maximum choice in each category is 20 (i.e. the number of parents that have chosen the specific word). Shaded cells identify the signal with highest frequency for each mushroom or cluster.

action avoid/approach. In the 6-unit cluster, four nodes are used to name the three types of edible mushroom and one for the category of toadstools. These signals are thus proto-linguistic nouns. The use of the terms ‘‘verb’’ and ‘‘noun’’ here must not be construed against the nouns and verbs of human languages, but rather as belonging to a compositional language with predicate–argument structure.

2.4. Cultural variation and the Baldwinian evolution of language

The simultaneous use of a genetic algorithm to model biological evolution and a back-propagation learning algorithm to model the cultural transmission and evolution of language, permits study of interaction between phylogenetic and ontogenetic mechanisms in language evolution. Some theories of language evolution explicitly refer to such phenomena, such us hypothesis on the Baldwinian evolution of the Language Acquisition Device (Pinker and Bloom, 1990; Briscoe, 2000). The Baldwin effect (Baldwin, 1896; Turney et al., 1996) consists in the fact that a trait or behavior, which every member of a species learns in ontogenesis, becomes part of the genetic makeup of that species in evolutionary time. In other words, genetic assimilation ensures that what initially needs to be learnt eventually becomes innate. The application of the Baldwin effect to the Language Acquisition Device hypothesizes that linguistic knowledge that was ontogenetically acquired by our first language speaking ancestors has gradually been assimilated in our genotype. This would explain the innate capacity of human children to process linguistic knowledge, as proposed by language nativists such as Chomsky (1975).

In language cultural variation – changes in the form of words (lexicons) and/or syntax – has been hypothesized to use interaction between evolution and learning. The present model can simulate how cultural variation and Baldwinian evolution interact in the emergence of compositional languages. The languages in this model can be subject to the process of cultural variation that allows languages to change and evolve over generations. Cultural variation and language distortions occur at the interface between old and new generations of language users. This is implemented through the addition of a random value between ± 0.5 to the parents' output linguistic nodes. The noise can produce a stochastic effect by randomly changing some of the winner-takes-all nodes in the output linguistic clusters. Such a mechanism introduces a variation in the language environment which is needed in experiments looking for a Baldwin effect. The use of noise in language learning is motivated by the fact that human languages are not transmitted in a noise-free fashion.

To investigate the Baldwin effect with, and without noise, we analyzed the language learning error curve at the first (gen. 301) and last (gen. 400) language learning generation. The expectation is that the presence of a Baldwin effect will show a low and flat error curve since the last generation. That is, agents that have assimilated the language to their genotype will be born, at generation 400, with a low error rate due to their good knowledge of the language. Fig. 2 reports error curves for the simulation without cultural variation (2a) and with learning noise (2b). A clear Baldwin effect is found where there is no cultural variation. At generation 400, since the first epoch's error is already low, agents are born with good lexical knowledge. In the case of cultural variation, no assimilation of the lexicon has happened, since the initial learning error at generation 400 is the same as that of generation 301. However, the learning curve shows a sudden decrease of language errors in the first 5 epochs. Detailed analyses of the mechanisms behind a steep learning curve (Munroe and Cangelosi, 2003) show a different type of Baldwin effect in this simulation.

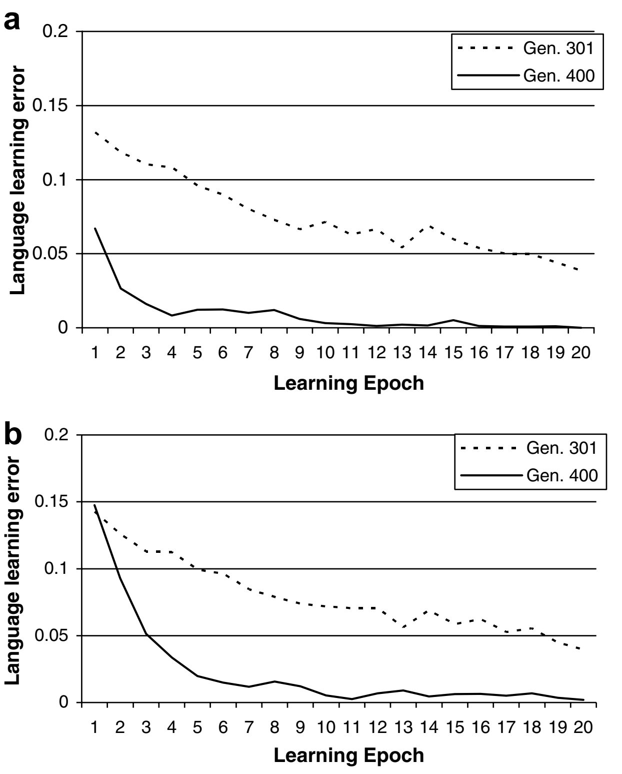

Fig. 2. Language learning curves in the simulation with cultural variation (2a) and without (2b). The dotted line is the language error curve at the first generation when language is introduced (gen. 301) and the continuous line at the last generation (gen. 4000). (a) Without cultural variation. (b) With cultural variation.

2.5. The sensorimotor bases of shared compositional languages

The above simulations on the evolution of shared signaling shows how evolution can unite external and internal aspects of symbol grounding to form-based categorization (phrase structure) or, for example, auditory and motor patterns of speech. The same modeling methodology can be applied to other aspects of language, such as the interaction between language and action (Pecher and Zwaan, 2005). What happens if, instead of focusing on perception, simulated agents have an action component? How does action knowledge interact with linguistic categories?

2.6. The model and simulation setup

This simulation follows on the foraging modeling scenario reported both above and in other language evolution simulations (Cangelosi, 2001). In this case, however, foraging behavior is replaced by an object manipulation task. Use of a robotic arm allows for a richer action repertoire based on using different objects. While this implementation exploits two actions (push/pull) and two objects (vertical and horizontal bars), the model can be scaled up to include larger numbers of action and object. The current task allows us to study the evolution of a compositional lexicon based on the proto-syntactic categories of nouns such as names of objects, verbs or names of actions. The 2×2 lexicon (push, pull and object_A, object_B) is provided to the agents as linguistic input from the beginning and, as shown by the Balwinian evolution study, does not lead to emergence.

The adaptive agent model uses an object manipulation scenario with an individual agent whose 2-segment arm (Fig. 3) has two degrees of freedom. The agent has a retina for perceiving the environment where one of two objects (object $A =$ vertical bar, object $B =$ horizontal bar) is always present in a range of different positions. As the environment is perceived as a 25-pixel grid (5×5 square grid) the arm, which is initially positioned on the object, can either move the object towards the shoulder (pull) or push it away. Apart

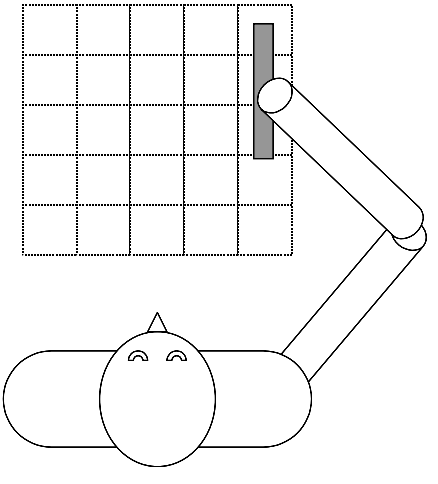

Fig. 3. The agent and its environment.

from retina input, the agent receives proprioceptive information about arm position and instructions using verbs and nouns.

During the organism's lifetime, the agent may be subject to various tasks. The first task is called *No-Language* and consists of a pure perceptual/categorical decision. In this case, the agent processes information from the retina that serves to decide what to do with the object identified. By default, in absence of linguistic instructions, object A is pushed, and B pulled. Ten more tasks that involve linguistic instructions are also available. In five of these, both retinal and linguistic information serve as input. In the remaining five tasks, only a linguistic instruction is provided and retinal units are set to 0. While proprioceptive input is present in all 11 tasks, the five linguistic instructions are shown in Table 2.

The behavior of the agent is controlled by a four-layer feed-forward neural network (Fig. 4). The input layer contains 25 retina units (the 25 cells of the visual grid), 4 proprioceptive units (for the position of the two pairs of muscles), and 4 linguistic units (these encode the two nouns and two verbs with a localist/binaty code). The first hidden layer consists of two separate sets of units, 5 that preprocess proprioceptive information, and

Table 2
Five linguistic conditions

Linguistic condition	Object present A	Object present B
Noun_Only	"A"	"B"
Default_Verb_Only	"PUSH"	"PULL"
Opposite_Verb_Only	"PULL"	"PUSH"
Default_Verb + Noun	"PUSH A"	"PULL B"
Opposite_Verb + Noun	"PUSH A"	"PULL B"

They are repeated twice, once while the agent sees the object and receives the retina input, and once in absence of visual input.

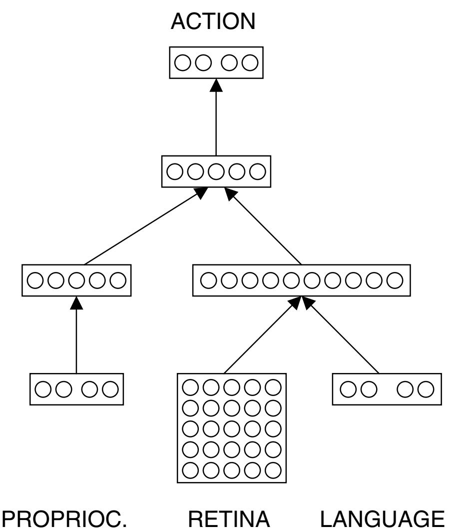

Fig. 4. Neural controller. The first hidden layers perform sensory processing (separately for the motor and for the vision and language input), whilst the second hidden layer has a sensorimotor integration role.

Please cite this article in press as: Cangelosi, A., Adaptive agent modeling of distributed ..., Lang. Sci. (2007), doi:10.1016/j.langsci.2006.12.026

10 that preprocess its visual and linguistic counterpart. The second hidden layer has 5 units for integration of sensory (retina + language) and motor (proprioception) information. The output layer contains four motor units that control the extension/contraction of four arm muscles (a pair of extension/contraction muscles per arm segment). The output activation simulates the force applied to each muscle.

The modular architecture of this network, with its separate organization of intermediate layers creates two independent preprocessing modules that deal with both the first hidden layer's sensorial and the motor information and the subsequent integration of all modalities in a second hidden layer.

2.7. Results on the evolution of nouns and verbs

A population of 80 agents simulated the evolution of sensorimotor and linguistic behavior. During the first part of the simulation (1000 generations), the agents evolved only the sensorimotor ability to identify and act on an object. This enabled us to pre-evolve a population of agents that respond appropriately respond to the objects before language is introduced. In the next 1000 generations, both sensorimotor behavior and the linguistic ability to comprehend and respond to instructions are made subject to selective selection and evolution.

A genetic algorithm is used to evolve genotypes associated with the agent's neural controller. At the beginning of the simulation, 80 genotypes are randomly initialized with connection weights in the range of ± 1. Each consists of 405 weights, encoded as real numbers. In the first stage sensorimotor stage of evolution, a generation consists of 18 subtasks (2 objects $\times$ 9 positions) in the *No_Language* condition. Each of these lasts for the 20 input/output cycles needed to move an object from its initial position to the target. In the remaining 1000 generations, 198 subtasks are executed for each agent. These include 18 *No Language* subtasks and 180 (18×10) linguistic subtasks. Each individual agent experiences the tasks in a random order. A fitness formula is used to assess the agents' levels of adaptation and to compute the total number of subtasks successfully completed by each individual in all conditions. A subtask is successful when the object is pulled to a distance of five points or less from the agent's shoulder (or pushed to a distance of 45 points or more from the agent's shoulder). In the linguistic tasks with verbs, verbal input determines, regardless of the objects default action, the correct push/pull action. When no verbs serve as input, the correct behavior always corresponds to the default action associated with each of the two objects.

At the end of each generation, fitness points are used to rank the agents. The 20 with highest levels of fitness are then are selected for reproduction. Each individual asexually generates four offspring with the same genotype as its single parent. The new genotypes are then subject to random mutations where 5% of the weights are changed by adding a random quantity in the range ± 1.

A set of 20 replications was used to test the model. In each of these, an initial population with different random weights was used. Overall, in all 20 populations, agents evolved the ability to act on objects and respond to linguistic instructions. In the final generation, the average number of subtasks successfully executed is 17.6 for the *No_Language* task and (17.96 in chart of Fig. 5) for its linguistic equivalent. This simulation generally confirms the behavioral and fitness results observed in other studies with a similar model (e.g. Cangelosi and Parisi, 2001).

Please cite this article in press as: Cangelosi, A., Adaptive agent modeling of distributed ..., Lang. Sci. (2007), doi:10.1016/j.langsci.2006.12.026

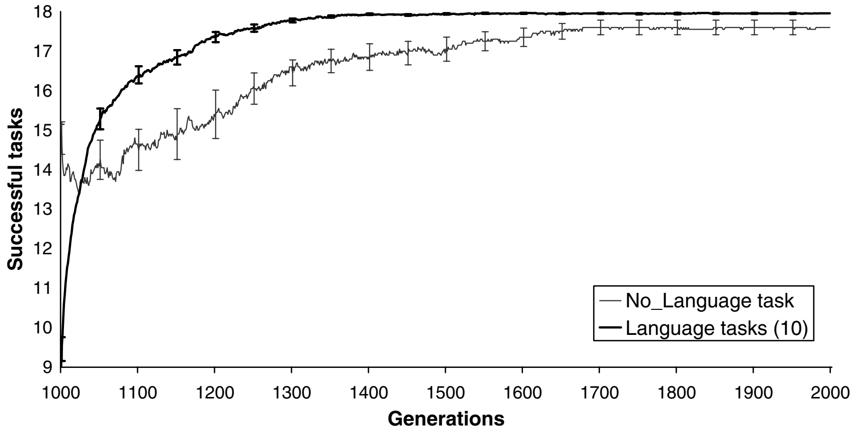

Fig. 5. Average fitness and standard error values for the single *No_Language* tasks and the 10 linguistic tasks. Data averaged over the fitness values of the single best individuals of the 20 replications.

2.8. Analyses of sensorimotor and linguistic representations

The use of a modular architecture for the agent's neural controller had the function of separating the sensory processing of proprioceptive and visual input in the early stages of information processing (first hidden layer) from subsequent integration of sensorimotor knowledge (in the second hidden layer). In order not to bias the integration of linguistic and sensorimotor information, the linguistic input was directly connected to both hidden layers. In this section we will analyze the functional organization on the neural network to identify the level and mode of integration of the linguistic, perceptual and motor knowledge.

The analyses are based on the categorical perception methodology. Categorical perception (Harnad, 1987) refers to the ability of individuals to perceptually categorize input stimuli by warping the similarity space of internal categorical representations. The warping phenomenon results in the compression of within-category distances between members of a category, and the expansion of between-category distances amongst members of different categories. In psychological experiments on category leaning, the distances, i.e. differences, between category members can be measured through multidimensional scaling methods (Andrews et al., 1998; Goldstone, 1994). In neural network models of categorization, Euclidean distances in the multidimensional hidden activation space can accurately measure these 'warping' effects (Cangelosi and Harnad, 2000; Cangelosi et al., 2000; Tijsseling and Harnad, 1997).

In this study, the agents build categorical representations on the environment (i.e. the two objects) through neural controllers. Such categorization is achieved both though perceptual analyses, as in the *No_Language* task, and through linguistic input, as in the 10 linguistic instruction tasks. The Euclidean categorical perception data were computed separately for the two intermediate layers (visual sensory processing in the first hidden layer and sensorimotor integration in the second layer). No analyses are conducted with proprioceptive sensory processing units because they only process motor feedback information. Data were computed and averaged using the hidden unit activations of the best agent in

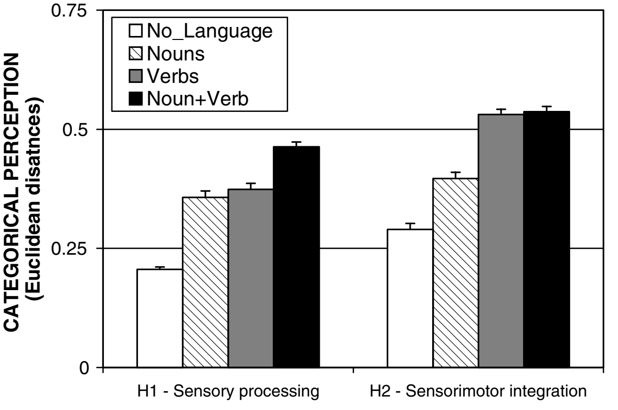

Fig. 6. Data on the categorical perception analyses (Euclidean distances) of verb and noun processing in the two hidden layers (neural architecture of Fig. 2). The *Y*-axis is the average Euclidean distances of the best 20 individuals (one per population). Error bars refer to the standard error.

each of the 20 replications. To compute the between-category distances, the sets of N hidden activation values h for the object A ($h1_A$, $h2_A \ldots hn_A$) and B ($h1_B$, $h2_B \ldots hn_B$) are recorded. These constitute the coordinates of two points in a N-dimensional Cartesian plane. Subsequently, the Pythagoras theorem is applied to obtain the Euclidean distance d_{A-B} between the two points:

$$d_{A-B} = \sqrt{(h1_A - h1_B)^2 + (h2_A - h2_B)^2 + \cdots + (hn_A - hn_B)^2}$$

Fig. 6 reports the average Euclidean distances of four groups of tasks: *No_Language* (white column), Noun-only tasks (striped column) and Verb-only tasks (dark gray column) and Verb + Noun tasks (black column). Results indicate that, in both hidden layers, nouns produce enhanced categorical perception effects in comparison with the baseline *No_Language* task. This is consistent with literature data on how language affects category learning (e.g. Cangelosi and Harnad, 2000). Verbs differ from nouns only in the second hidden layer where a motor (i.e. proprioceptive) signal is integrated with sensory (i.e. visual) input. The forth column shows that when verbs and nouns are combined, they produce the enhanced categorical perception effects typical of verbs (cf. similar height of the gray and black columns in the second hidden layer). The fact that the extent of the warping effects in the first hidden layer is similar in both nouns and verbs indicates that at this stage of processing the two word categories are undifferentiated. In the same layer of hidden units, the significant difference between the two Verb_only/Noun_only conditions and that of the Verb + Noun condition, also suggests that verbs do not play their typical role at this level of processing. Overall, the data support the hypothesis that verbs produce enhanced warping effects solely where hidden units play a direct role in integrating motor and sensory information that leads to motor response.

These results are consistent with other investigations of the agents' internal representations using synthetic brain imaging methodology (Cangelosi and Parisi, 2004; Cangelosi, 2004). The analyses demonstrate that when the agents are processing verbs the sensorimotor integration areas of the network are more active while, by contrast, sensory units are more active during noun processing.

3. Conclusions

In this paper we present a computational approach, based on adaptive agent simulations, to the study of language. This focuses on the integration of internal and external factors in language acquisition and evolution. The two core assumptions are that language is directly grounded into the agent's own cognitive representations (internal symbol grounding) and into the social and physical environment in which they develop their linguistic capabilities (external symbol grounding).

In the model on the emergence of syntax (Section 2) we demonstrated the importance of social and cultural factors in the evolution of shared lexicons. Although the physical entities in the environment provide a relatively stable set of entities upon which the lexicon can be built, social factors, and in particular cultural variation during language learning, allow the agents to build different linguistic representations. This is especially important in the earlier stages of language learning when agents can be born with different linguistic knowledge. When cultural variation is allowed, they have a predisposition for rapid acquisition of the specific lexicon to which they are exposed. Instead, when no cultural variation is permitted, agents are born with a good knowledge of the lexicon used by their parents.

The models assume a 'double digitality' in that, from the first, agents identify distinct signals *and* kinds. In examining the compositional lexicon (viz. the processing of signals associated with different functions) agents use robotic arms to push and pull vertical bars. This brings out that there is interaction between linguistic and physically grounded information. Measures of categorization show (a) signals influence processing, i.e. are integrated with real-time activity and cognitive dynamics; (b) whereas non-functors only affect the sensory processing level, verbs enhance sensorimotor integration, so folk distinctions may have a functional basis; (d) the categories rely on processing stages or time domains.

In the model on the evolution of verbs and nouns for action manipulation tasks, the focus is on the sensorimotor grounding of language and its neural representation. The categorical perception analyses show that the names of actions (verbs) produce enhanced warping effects solely where hidden units play a direct role prepare motor response by integrating the motor and sensory information.

The view of language proposed by these simulation studies is that of a distributed linguistic system strongly dependent on the coupling of embodiment factors, internal representation properties and social factors. The embodiment factors refer to the physical and perceptual properties of both entities in the world and the sensorimotor mediation that the agent's embodiment produces. The agents' perceptual system and its motor actuators affect the type of interaction with the external objects and the type of lexicon needed to represent it. For example, in the lexicon shown in Table 1, the agents do not build differentiated lexical entries for the three toadstools because their sensorimotor apparatus treats them as equal (all poisonous mushrooms need to be avoided the same way). The contribution of internal representation factors refers to the properties of internal, neural representations that closely match the organism embodiment constraints. Representations for verbs and nouns differ depending on the functional role (sensory processing vs. sensorimotor integration) of the neural substrate responsible for their processing. Finally, the social factors, such as cultural variation, affect the type of language that the agents possess during the various evolutionary and developmental stages.

Please cite this article in press as: Cangelosi, A., Adaptive agent modeling of distributed ..., Lang. Sci. (2007), doi:10.1016/j.langsci.2006.12.026

The adaptive agent approach proposed here has great potential for further studies on the distributed structures of language. For example, future studies could look at the role of artifacts in the evolution and acquisition of language. In a related adaptive agent model of the evolution of artifacts (Ugolini and Parisi, 1999), agents are able to create artifacts by modifying the external environment and increasing their reproductive chances. Artifacts can be culturally transmitted as each generation reproduces the artifacts of the preceding generation. If the reproduction of artifacts is selective and new variants are added in each generation, artifacts evolve. This model could, for example, be coupled with one of the foraging models of language emergence to establish how artifacts might contribute to the evolution of language. It could also be extended to the evolution of writing.

The proposed modeling methodology has important implications for the distributed view of language discussed elsewhere in this issue. First, it is consistent with the view that languages are not codes in the sense attacked by Love, Carr and Kravchenko. Utterances are physical events based in a history of interactions that do not rely on the denotations of abstract linguistic units. In the simulations described above, linguistic categories draw on a history of interaction of a population of agents with a social and physical environment. Investigation of physical social symbol grounding suggests that acts of utterance relate to action through something like Menary's (this issue) 'cognitive integration'. As argued by several contributors to this issue, relations to the environment are fundamental. In the external or social domain agents realize values (Hodges, this issue) or develop stable signals (Ross, this issue) that, given diverse inner representations, depend on how interaction warps the similarly space. The simulations also fit Linell's (this issue) view that dialogical signaling (i.e. using linguistic artifacts), constitutes first-order language that is sustained by both social practices and the representations of a dialogue ready brain. This is consistent with the internal and external symbol grounding views according to which language is grounded both inwards and outwards. Finally, the model fits Kravchenko's (this issue) view that the key to complex behavior lies in an agent's self-organizing activity while, at the same time, it also depends on the agents' use of independent features and, especially, conventions.

Acknowledgements

The research presented here was supported by various grants from the UK Engineering and Physical Sciences Research Council, The Nuffield Foundation, and the European Office of Aerospace Research and Development (US Air Force Laboratory). I thank Stephen Cowley for his useful comments on the paper and the discussion of the close relationship between the language modeling methodology presented here and the various theoretical positions of the Distributed Language community.

References

Andrews, J., Livingston, K., Harnad, S., 1998. Categorical perception effects induced by category learning. Journal of Experimental Psychology: Learning, Memory, and Cognition 24, 732–753.

Baldwin, J.M., 1896. A new factor in evolution. The American Naturalist 30, 441–451.

Barsalou, L., 1999. Perceptual symbol systems. Behavioral and Brain Sciences 22, 577–609.

Briscoe, E.J., 2000. Grammatical acquisition: Inductive bias and coevolution of language and the language acquisition device. Language 76, 245–296.

Burgess, C., Lund, K., 1997. Modeling parsing constraints with high-dimensional context space. Language and Cognitive Processes 12, 177–210.

Cangelosi, A., 2001. Evolution of communication and language using signals, symbols and words. IEEE Transactions on Evolutionary Computation 5 (2), 93–101.

Cangelosi, A., 2004. The sensorimotor bases of linguistic structure: experiments with grounded adaptive agents. In: Schaal, S., Ijspeert, A.J., Billard, A., Vijayakumar, S., Hallam, J., Meyer, J.-A. (Eds.), . In: Proceedings of the Eighth International Conference on the Simulation of Adaptive Behaviour: From Animals to Animals, vol. 8. MIT Press, Cambridge MA, pp. 487–496.

Cangelosi, A., 2005. Grounding symbols in perceptual and sensorimotor categories: connectionist and embodied approaches. In: Cohen, H., Lefebvre, C. (Eds.), Handbook of Categorization in Cognitive Science. Elsevier, Oxford.

Cangelosi, A., 2006. The grounding and sharing of symbols. Pragmatics and Cognition 14 (2), 275–285.

Cangelosi, A., Greco, A., Harnad, S., 2000. From robotic toil to symbolic theft: grounding transfer from entry-level to higher-level categories. Connection Science 12 (2), 143–162.

Cangelosi, A., Harnad, S., 2000. The adaptive advantage of symbolic theft over sensorimotor toil: grounding language in perceptual categories. Evolution of Communication 4 (1), 117–142.

Cangelosi, A., Parisi, D., 2001. How nouns and verbs differentially affect the behavior of artificial organisms. In: Moore, J.D., Stenning, K. (Eds.), Proceedings of the 23rd Annual Conference of the Cognitive Science Society. Lawrence Erlbaum, London, pp. 170–175.

Cangelosi, A., Parisi, D., 2004. The processing of verbs and nouns in neural networks: insights from synthetic brain imaging. Brain and Language 89 (2), 401–408.

Cangelosi, A., Riga, T., 2006. An embodied model for sensorimotor grounding and grounding transfer: experiments with epigenetic robots. Cognitive Science 30 (4), 673–689.

Chomsky, N., 1975. Reflections on Language. Pantheon Press, New York.

Cowley, S.J., 2005. A distributed view of language origins. In: Proceedings of Second International Symposium on the Emergence and Evolution of Linguistic Communication (EELC'05). AISB Press, London.

Glenberg, A., Kaschak, M., 2002. Grounding language in action. Psychonomic Bulletin and Review 9 (3), 558–565.

Goldstone, R., 1994. Influences of categorization of perceptual discrimination. Journal of Experimental Psychology: General 123, 178–200.

Harnad, S., 1990. The symbol grounding problem. Physica D 42, 335–346.

Harnad, S. (Ed.), 1987. Categorical Perception: The Groundwork of Cognition. Cambridge University Press, New York.

Harnad, S., 1993. Grounding symbols in the analog world with neural nets. Think 2 (1), 12–78.

Hsiao, K., Mavridis, N., and Roy, D., 2003. Coupling perception and simulation: steps towards conversational robotics. In: Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems, Las Vegas, 2003.

Hazlehurst, B., Hutchins, E., 1998. The emergence of propositions from the co-ordination of talk and action in a shared world. Language and Cognitive Processes 13, 373–424.

Holland, J.H., 1975. Adaptation in Natural and Artificial System. The University of Michigan Press, Ann Arbour, MI.

Fodor, J.A., 1975. The Language of Thought. Harvard University Press, Cambridge MA.

Marocco, D., Cangelosi, A., Nolfi, S., 2003. The emergence of communication is evolutionary robots. Philosophical Transactions of the Royal Society London A 361, 2397–2421.

Miikkulainen, R., 1994. Integrated connectionist models: building Ai systems on subsymbolic foundations. In: Honavar, V., Uhr, L. (Eds.), Symbol Processors and Connectionist Network Models in Artificial Intelligence and Cognitive Modelling: Steps Toward Principled Integration. Academic Press, pp. 483–508.

Munroe, S., Cangelosi, A., 2003. Learning and the evolution of language: the role of cultural variation and learning cost in the Baldwin Effect. Artificial Life 8, 311–339.

Pecher, D., Zwaan, R.A. (Eds.), 2005. Grounding Cognition: The Role of Perception and Action in Memory, Language, and Thinking. Cambridge University Press, Cambridge, UK.

Pinker, S., Bloom, P., 1990. Natural language and natural selection. Behavioral and Brain Science 13, 707–784.

Roy, D., 2005. Semiotic schemas: a framework for grounding language in action and perception. Artificial Intelligence 167, 170–205.

Rumelhart, D.E., Hinton, G., Williams, R., 1986. Learning internal representations through error propagation. In: Rumelhart, D.E., McClelland, J.L., The PDP Research Group (Eds.), . In: Parallel Distributed Processing: Experiments in the Microstructure of Cognition, vol. 1. MIT Press, Cambridge MA.

Please cite this article in press as: Cangelosi, A., Adaptive agent modeling of distributed ..., Lang. Sci. (2007), doi:10.1016/j.langsci.2006.12.026

Steels, L., Kaplan, F., 1999. Collective learning and semiotic dynamics. In: Floreano, D., Nicoud, J., Mondada, F. (Eds.), Proceedings of ECAL99 European Conference on Artificial Life. Springer-Verlag, Berlin, pp. 79–688.

Steels, L., 2002. Grounding symbols through evolutionary language games. In: Cangelosi, A., Parisi, D. (Eds.), Simulating the Evolution of Language. Springer-Verlag, London.

Steels, L., 1999. In: The Talking Heads Experiment, vol. 1. Words and Meanings, Antwerpen.

Sun, R., 2002. Connectionist implementationalism and hybrid systems. Encyclopedia of Cognitive Science. Nature Publishing Group, MacMillan.

Tijsseling, A., Harnad, S., 1997. Warping similarity space in category learning by backprop nets. In: Ramscar, M., Cambouropolos, E., Pain, H. (Eds.), Proceedings of SimCat 1997: Interdisciplinary Workshop on Similarity and Categorization. Edinburgh University, pp. 263–269.

Turney, P., Whitley, D., Anderson, R.W., 1996. Evolution, learning, and instinct: 100 years of the Baldwin effect. Evolutionary Computation, 4(3) pages (Special issue on the Baldwin effect).

Ugolini, M., Parisi, D., 1999. Simulating the evolution of artifacts. In: Floreano, D., Nicoud, J.D., Mondada, F. (Eds.), Advances in Artificial Life. ECAL99. Springer, New York.

Vogt, P., 2002. The physical symbol grounding problem. Cognitive Systems Research 3 (3), 429–457.

Zentall, T.R., Jackson-Smith, P., Jagielo, J.A., Nallan, G.B., 1986. Categorical shape and color coding by pigeons. Journal of Experimental Psychology: Animal Behavior Processes 12 (2), 153–159.