

# Establishing Communication Systems without Explicit Meaning Transmission

Andrew D.M. Smith

Language Evolution and Computation Research Unit,  
Department of Theoretical and Applied Linguistics, University of Edinburgh, UK  
andrew@ling.ed.ac.uk

**Abstract.** This paper investigates the development of experience-based meaning creation and explores the problem of establishing successful communication systems in a population of agents. The aim of the work is to investigate how such systems can develop, without reliance on phenomena not found in actual human language learning, such as the explicit transmission of meaning or the provision of reliable error feedback to guide learning. Agents develop individual, distinct meaning structures, and although they can communicate despite this, communicative success is closely related to the proportion of shared lexicalised meaning, and the communicative systems have a large degree of redundant synonymy.

## 1 Introduction

There is a growing body of literature in which investigations into the evolution of language<sup>1</sup> are carried out by computer simulation [8,1,4]. For most of these researchers, the evolution of language is regarded as essentially being equivalent to the evolution of syntax, because the use of syntactic structure is seen as the main difference between animal and human communication systems. For example, vervet monkeys have a well-known communication system which allows them to distinguish different predators [5], but they do not combine their signals to convey complex meanings. Kirby [9] has shown that the simple ability to create general rules, by taking advantage of coincidental correspondences between parts of utterances and parts of meanings, can result in the emergence of syntax, as general rules generate more utterances than idiosyncratic rules, and are therefore replicated in greater numbers in following generations. Similar accounts [2,10] also show syntax emerging as a consequence of the recognition and coding of regularities between signals and meanings.

Nehaniv [14] has pointed out, however, that syntax only develops successfully from unstructured signals because the signals are coupled with meanings which are already structured, and it is no coincidence that the emergent syntactic structure parallels the pre-existing semantic structure. In these simulations,

---

<sup>1</sup> This field is concerned not with the evolution of particular languages, such as English, from their ancestor languages, but rather with the general capacity, apparently unique to humans, for using infinitely expressive communication systems [13].

the meanings are also explicitly part of the linguistic transfer from speaker to hearer, therefore obviating the critical problem, exemplified by Quine [17](p. 29–30), of how a learner determines the meaning which a signal intends to convey. Furthermore, attempts to develop learnt communication systems frequently involve some sort of reinforcement learning process [20,6], which has the primary role in guiding the learning mechanism. Oliphant [15] points out, however, that such error signals, which work well on an evolutionary timescale, are less useful over an individual’s lifetime where failure might mean immediate death, and indeed even the very existence of reliable error signals is questioned by many authors on child language acquisition [3].

If we try to define the meaning of a word, we find ourselves caught in a kind of lexical web, where words can only be defined by their relationship to other words, and in terms of other words. There is no obvious way of entering this web, unless at least some words are grounded in reality [7], such that they can be used to point out actions and objects in the real world. It is reasonably uncontroversial to say that meanings must capture patterns of categorisation (whether categories are defined in classical terms of shared features or prototypes [21]) which enable us to state, for instance, which things are *rabbits* and which are not. Furthermore, meanings are not innate, but are created anew in each language learner, who creates an individual system of meaning based on their experiences [3].

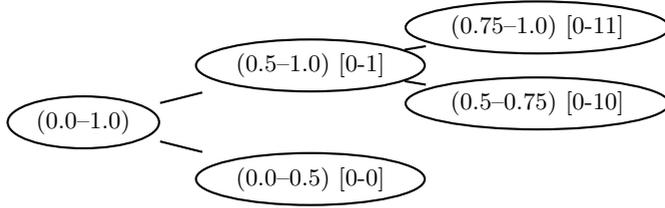
Our aim is to model, in a population of agents, the creation of meanings by explicit categorisation, and then to investigate the spread of meanings through the population, without the meanings themselves being transferred between agents, and without any error signals to reinforce the learning process.

## 2 Meaning Creation by Object Discrimination

In order to develop a model of independent, grounded meaning creation, we establish a simple world of agents and objects, similar to that described by Steels [19], in which the objects can be described in terms of their features<sup>2</sup>, which are intrinsically meaningless, but which can be thought of in terms of more imaginable language-like features such as *colour*, *height* or *smell*. The agents in the model world interact with the objects by using *sensory channels*, which are sensitive to the corresponding features of objects, and can detect whether a particular value of a feature falls between two bounds. Initially, the channels can only detect that a value falls between 0.0 and 1.0, but the agents have the power to split the sensitivity range of a channel into two discrete segments, resulting in a *discrimination tree* [20]. The nodes of a discrimination tree can be considered categories or *meanings*,<sup>3</sup> as seen in the sensory channel in figure 1, which has been refined twice, and has four new meanings.

<sup>2</sup> Feature values are represented as pseudo-randomly generated real numbers which are normalised to lie between 0.0 and 1.0

<sup>3</sup> Meanings are given in the notation *sc-path*, where *sc* identifies the sensory channel, and *path* traces the path from the tree root to the node in question, where 0 signifies a lower branch and 1 an upper branch.



**Fig. 1.** A discrimination tree (channel 0) which has been refined twice. Each node shows the bounds between which it is sensitive, and the meaning to which it corresponds (following Steels).

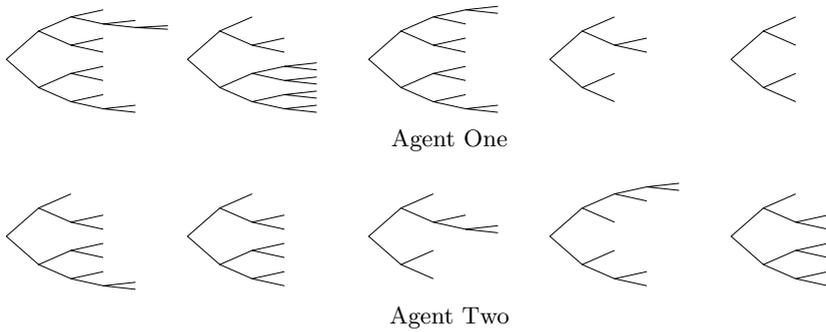
In order to provide a framework for the unguided refinement of the sensory channels based on observation, we follow Steels [19] in using discrimination games, in which an agent attempts to distinguish one object from a larger set of objects. Each game proceeds as follows:

1. An agent considers a random set of objects (the *context*), one of which is chosen at random to be distinguished from the others and is called the *topic*.
2. The agent investigates all its sensory channels to categorise the objects.
3. If the topic is uniquely identified by any category, the game succeeds.
4. If the game fails, the agent refines a randomly-chosen sensory channel.

Object	Categories/Meanings		
A	0-0	1-00	2-111
B	0-11	1-1	2-110
C	0-0	1-1	2-111
D	0-10	1-01	2-10

The above table shows an agent categorising objects as part of a discrimination game. The agent has four objects A-D, and has categorised them with three sensory channels. If the aim of this game is to discriminate *B* from the context *ACD*, then the game can succeed, as both 0 – 11 and 2 – 110 are possible distinguishing categories. On the other hand, if the aim is to distinguish *C* from the context *ABD*, then the game will fail, as none of the categories which *C* falls into distinguish it from all the other objects. Failure triggers the refinement of a random channel, creating more detailed categories, which *may* be useful in future games. Over time, the agents develop their sensory channels such that the discrimination games nearly always succeed, though the extent to which an individual channel is refined depends on the number of channels which the agent has: the more channels, the fewer refinements on each are necessary.

Figure 2 shows the idiosyncratic meaning representations of two agents in the same world. The first agent has developed the first three channels to a greater extent than the second agent, who in turn has developed the fourth and fifth channels more extensively. It is helpful to quantify the amount of difference between two trees  $t_1$  and  $t_2$ , which we can do by averaging the proportion of



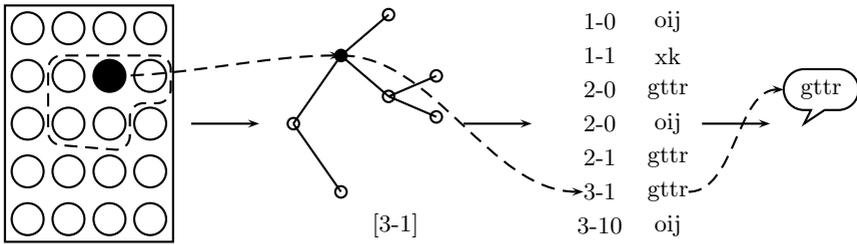
**Fig. 2.** Two agents each have five sensory channels, with which they construct different representations of the same world.

nodes in tree  $t_1$  which are also in  $t_2$ , and the proportion in  $t_2$  which are also in  $t_1$ . Averaging over all the trees in figure 2, the two meaning representations have a *meaning similarity* measure of 75%. It is important to note that both agents are successful in the discrimination games, and so their representations are equally good descriptions of their world. This model, then, satisfies one of our goals, namely that the agents are not given innate meanings, but can create inventories of basic concepts individually, based on their own experiences.

### 3 Communication

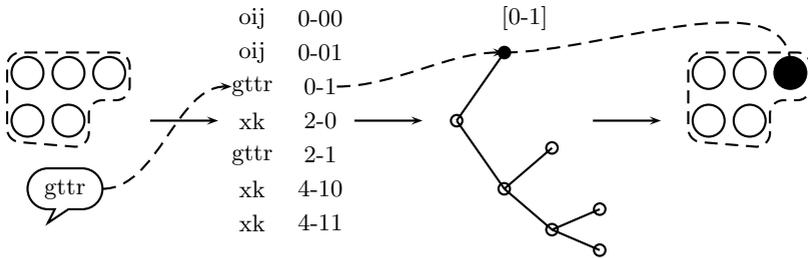
The next step is to investigate whether the agents can communicate with each other, using the meanings they have constructed. Clearly the agents must be able to use some sort of signals, and so they are endowed with the ability to create signals from random strings of letters, and to express and understand these signals without error. In addition, they maintain a dynamic lexicon of associations between signals and meanings, which develops as they participate in the experiments, and which they use in order to make decisions about their communicative behaviour. *Communicative success* occurs if the speaker and hearer are both referring to the same object, but it is not necessary for them to use the same meaning to do so.

A *communicative episode* is played between two agents chosen at random, the *speaker* and the *hearer*. Figure 3 shows a model of the speaker's role, which begins with a discrimination game, in which meanings which can distinguish the topic (filled circle) from the rest of the context (dashed area) are collated. One of these meanings is chosen at random and then looked up in the speaker's lexicon. If the speaker cannot find a word for the meaning it is trying to convey, then it creates a random string of letters and stores this in its lexicon with the required meaning. Having obtained a word to convey the meaning, the speaker utters the word, and the focus passes to the hearer, who receives the word, and can observe the context in which it was uttered, shown in figure 4.



**Fig. 3.** A communicative episode begins with an agent chosen at random to be the speaker, who finds a meaning to distinguish the topic from the context, and utters a word to convey this meaning.

The word is decoded via the hearer’s lexicon into a meaning, and the hearer then establishes which object in the context (if any) is uniquely identified by the meaning it has chosen. If the referent (object) identified by the hearer corresponds to the speaker’s original topic, then the communication episode succeeds. The success or failure of a communication game has no effect on the internalised representations of either agent. This model of communication conforms to our initial assumptions, as the internal meanings are explicitly *not* transmitted with the signals, and the agents do not receive feedback from each other about the success of their communicative or learning processes.



**Fig. 4.** The communicative episode continues with the hearer, who, given the context, decodes the word into a meaning which identifies an object.

## 4 The Lexicon

The mappings from meaning to signal and vice-versa are at the heart of the communication process, and are handled via a lexicon, which stores associations between meanings and signals, as well as a count of how often the signal-meaning

pair has been used (either uttered as a speaker or understood as a hearer), and a confidence probability, which represents the agent's confidence in the association between the signal and the meaning.

The confidence probability of the signal-meaning pair, consisting of signal  $s$  and meaning  $m$ , represents the history of all associations between words and meanings an agent has ever made, and is defined as the proportion of the times  $s$  has been used in which it has been associated with  $m$ , or  $\frac{Usage(s,m)}{\sum_l Usage(s,l)}$  where  $l$  is the number of entries in the lexicon. A short extract from an example lexicon is given below, only showing the entries for two of the signals (*gttr* and *oij*), and the meanings associated with them.

Signal	Meaning	Usage	Conf. Prob.
gttr	0-0	1	0.083
gttr	0-1	2	0.167
gttr	0-11	1	0.083
oij	1-0	9	0.600
gttr	2-0	4	0.333
oij	2-0	6	0.400
gttr	2-1	1	0.083
gttr	3-1	2	0.167
gttr	4-00	1	0.083

How does the speaker decide which signal to choose, when it is trying to express a particular meaning (say 2 – 0)? Given the lexicon above, the signal *oij* would seem a reasonable choice for two reasons: it has been associated with 2 – 0 on six occasions, compared to *gttr*'s four, and the agent is more confident in the association with *oij* (0.4) than *gttr* (0.33). However, Oliphant and Batali [16] have demonstrated an ideal strategy for achieving an accurate communication system, known as *obverter*, where the speaker chooses words which he knows the hearer will understand. Unfortunately, true obverter learning assumes that the speaker can read the lexicons of the other members of the population, to calculate the optimal signal to use for any meaning. Such mind-reading is not only unrealistic, but even avoids the need for communication at all, and so an alternative is needed. It seems reasonable to assume that the only lexicon the speaker has access to is its own, and so we assume that the speaker uses this as an approximation to that of the hearer. Instead of explicitly choosing the word that the hearer will understand, the speaker chooses the word that *it* would be most likely to understand if it was the hearer. Returning to the lexicon above, we can see that although *oij* has been associated with the meaning 2 – 0 on more occasions than *gttr*, if heard, it would actually be interpreted as 1 – 0 (because 1 – 0 is the meaning which maximises the confidence probability for *oij*), whereas *gttr* would be interpreted with the correct 2 – 0 meaning.

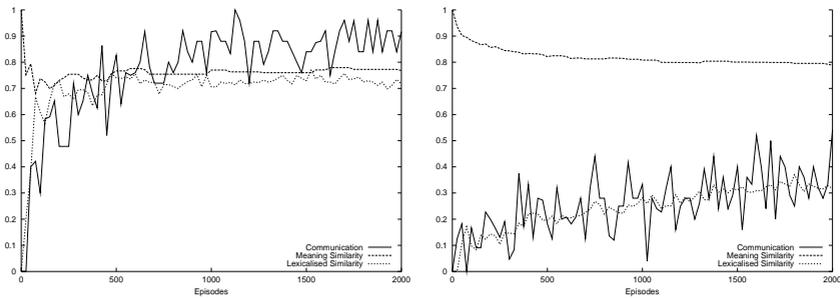
Interestingly, the agent would not find a word from its lexicon to express many meanings which do have some associations (e.g. 0 – 0, 3 – 1 etc.). One of the outcomes of obverter learning is the avoidance of ambiguity, so we find that, at any one time, each word in the lexicon is only used with one meaning,

although the particular meaning can of course change as the associations in the lexicon are updated. This means that, although there are eight meanings in the lexicon extract, only two of them are actually used by the speaker, and so only these can be regarded as being truly *lexicalised*.

We have seen how the speaker tries to second-guess the hearer and chooses words which are likely to be understood before uttering them, but a greater problem is faced by the hearer in understanding the meaning which is being conveyed. On hearing a signal, the hearer's only guide in determining the intended meaning is the observation of the context (which of course includes the target topic object). From this, the hearer constructs a list of all the possible meanings, that is, *all* meanings which categorise only one of the objects in the context. All these possible meanings are equally plausible, so the hearer associates each of them with the signal in its lexicon, adjusting its confidence probability for each accordingly. Over time, the interpretation of each word will tend to the speaker's intended meaning, if the two agents have identical meaning structures [18].

## 5 Results

The meaning structures constructed by the agents in our model world, however, are of course not only not identical, but also change over time. Under these circumstances, is it possible for the agents to communicate? Figure 5 (left) shows

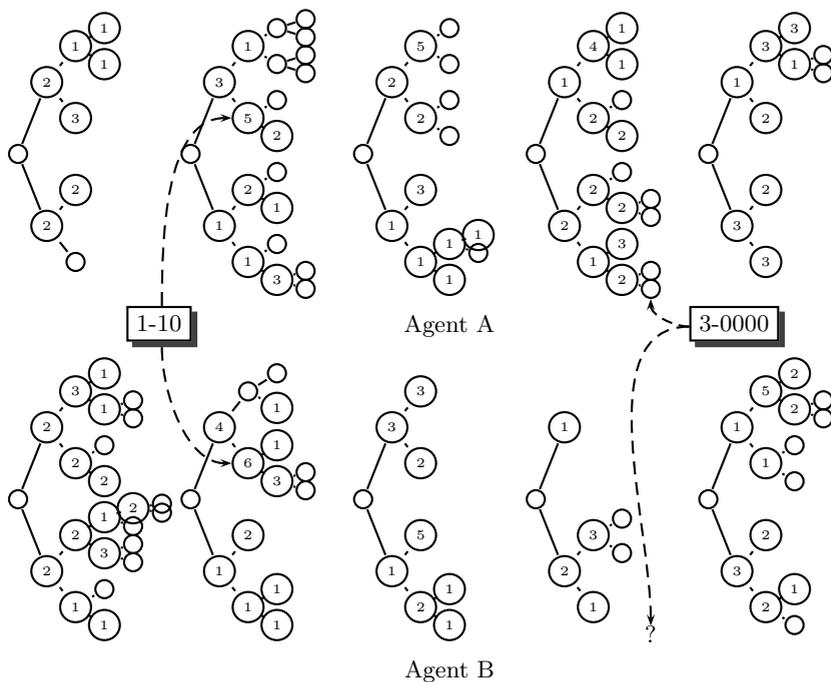


**Fig. 5.** Communicative success, meaning similarity, and lexicalised similarity for a population of two agents and 100 objects. Each discrimination game is played with a context size of five objects. The number of sensory channels available to each agent is five (left) and 100 (right).

that communication is successful a large percentage of the time, although it is not optimal, and does not appear to increase significantly after the initial rise to around 90%. The similarity of the agents' meaning structure drops initially, as the agents refine their sensory channels individually and separately, and then does not change significantly. This occurs because the pressure to develop meaning structure comes only from failure in discrimination games, and after an initial

flurry, the agents all have sufficiently detailed meanings to succeed in nearly all discrimination games. Once this state is achieved, the communication rate stops improving and remains fairly constant. If the number of sensory channels available is increased substantially (figure 5: right), a similar result is found, except that the rate at which communication stops improving is much lower. It can also be seen that the communication success rate is closely paralleled in both cases by the *lexicalised similarity* of the agents, which is defined in the same way as meaning similarity (see section 3), but only taking into account tree nodes which are lexicalised.

An interesting phenomenon which occurs in these kind of simulations is the large amount of synonymy which pertains in the lexicons, where more than one word is interpreted with the same meaning. As an example, after 1000 communicative episodes, two agents have the meaning structures shown in figure 6. Attached to each node on the discrimination trees is the number of words which



**Fig. 6.** Two agents each have five discrimination trees numbered 0-4. Each lexicalised node is marked with the number of words which would be interpreted as that meaning.

this agent would interpret as the meaning denoted by the node, or the number of *synonyms* attached to the meaning. For instance, we can see that there are five words which would each be interpreted by agent A as 1-10, and six which would

be interpreted as this by agent B. Further inspection (not shown) indicates that four of these synonyms have been lexicalised by both agents, suggesting a high level of redundancy, which is caused by *meaning drift*:

The interpretation of a word, of course, changes over time as the agents develop their experience of the word's use. Words are only created when an agent wants to express a meaning which isn't lexicalised. For example, in figure 6, agent A might wish to express the meaning 3 – 0000, but it does not have a word which it would interpret correctly, so it creates a new word *uj szo*. Agent B hears the new word, and creates a list of possible meanings. This list, however, cannot include A's meaning 3 – 0000, because B's meaning structure does not contain this meaning, and so B will lexicalise *uj szo* with a different meaning. Over time, B's preferred meaning is likely to be a more general meaning, which is shared by A.<sup>4</sup> There is now a difference of opinion over the meaning of *uj szo*, but crucially, agent A can continue to associate it with B's meaning, while B cannot associate it with A's original meaning. A's association between *uj szo* and the shared meaning gradually increases, until it eventually exceeds that of the original meaning. Both agents will now use *uj szo* for the more general meaning: the word's meaning has drifted. As a direct consequence, A no longer has a word with which it can express the meaning 3 – 0000. If it does need to convey this meaning, it must create another new word, and the cycle begins again.

Meaning drift is an inevitable characteristic of systems in which the agents' conceptual systems are not the same, if there are an unlimited number of signals, and there is little pressure to modify meaning structure. Inducing the meanings of words from context inevitably biases the meanings towards those meanings which are more general, and shared by the agents. Words which refer to specific meanings which are not shared will see their meanings drift to those which are shared, resulting in a large number of synonyms for the shared meanings, and few, if any, words at all for the agent-specific meanings.

## 6 Discussion

We have developed a world in which agents can communicate about their environment, without explicitly transferring meanings, without knowing exactly what the speaker is referring to, and without providing the learner with any feedback about communicative success, all criteria motivated by research into how human children acquire language [3]. Although communication can succeed in cases where agents refer to the same object with different meanings, the overall success of communication seems to be directly related to the amount of shared meaning structure in the agents. The communication system has a great deal of synonymy, caused by the differences in meaning structure and the unlimited number of possible signals. Work is under way to extend the model, focusing on ways to reduce synonymy, for instance by implementing the *principle of contrast* [12], and to investigate the effects of specific biases in meaning induction,

<sup>4</sup> Because general meanings are created before more specific meanings on the discrimination trees, they are more likely to occur in both agents' meaning structures.

such as the *shape bias* [11]. It is claimed that such biases explain the learning of meanings [3], and this work will go some way to showing where these claims are feasible.

## References

- [1] J. Batali. Innate biases and critical periods: Combining evolution and learning in the acquisition of syntax. In R. Brooks and P. Maes, editors, *Artificial Life 4: Proceedings of the Fourth International Workshop on the Synthesis and Simulation of Living Systems*, pages 160–171. Addison-Wesley, Redwood City, CA, 1994.
- [2] J. Batali. The negotiation and acquisition of recursive grammars as a result of competition among exemplars. In E. Briscoe, editor, *Linguistic Evolution through Language Acquisition: Formal and Computational Models*. Cambridge University Press, Cambridge, 2001.
- [3] Paul Bloom. *How Children Learn the Meanings of Words*. MIT Press, 2000.
- [4] E. Briscoe. Grammatical acquisition: Inductive bias and coevolution of language and the language acquisition device. *Language*, 76(2):245–296, 2000.
- [5] D. Cheney and R. Seyfarth. *How Monkeys See the World: Inside the Mind of Another Species*. University of Chicago Press, Chicago, IL, 1990.
- [6] Edwin D. de Jong. *Autonomous Formation of Concepts and Communication*. PhD thesis, Vrije Universiteit Brussel, 2000.
- [7] S. Harnad. The symbol grounding problem. *Physica*, D 42:335–346, 1990.
- [8] James R. Hurford. The evolution of critical period for language acquisition. *Cognition*, 40:159–201, 1991.
- [9] S. Kirby. Learning, bottlenecks and the evolution of recursive syntax. In Ted Briscoe, editor, *Linguistic Evolution through Language Acquisition: Formal and Computational Models*. Cambridge University Press, Cambridge, 2001.
- [10] V. Kvasnička and J. Pospíchal. An emergence of coordinated communication in populations of agents. *Artificial Life*, 5(4):319–342, 1999.
- [11] Barbara Landau, Linda B. Smith, and Susan S. Jones. The importance of shape in early lexical learning. *Cognitive Development*, 3:299–321, 1988.
- [12] E. M. Markman. *Categorization and naming in children*. MIT Press, 1989.
- [13] John Maynard Smith and Eörs Szathmáry. *The major transitions in evolution*. Oxford University Press, 1995.
- [14] Chrystopher L. Nehaniv. The making of meaning in societies: Semiotic and information-theoretic background to the evolution of communication. In *Proceedings of the AISB Symposium: Starting from Society — the application of social analogies to computational systems, 19–20 April 2000*, pages 73–84, 2000.
- [15] M. Oliphant. The learning barrier: Moving from innate to learned systems of communication. *Adaptive Behavior*, 7(3/4):371–384, 1999.
- [16] M. Oliphant and J. Batali. Learning and the emergence of coordinated communication. *Center for Research on Language Newsletter*, 11(1), 1997.
- [17] W. V. Quine. *Word and Object*. MIT Press, 1960.
- [18] A. D. M. Smith. The role of shared meaning structure in communication. Paper given at the 7<sup>th</sup> Human Behaviour & Evolution Society Conference, London, 2001.
- [19] L. Steels. Perceptually grounded meaning creation. In M. Tokoro, editor, *Proceedings of the international conference on multi-agent systems*. MIT Press, 1996.
- [20] L. Steels. *The Talking Heads Experiment*, volume I. Words and Meanings. Laboratorium, Antwerpen, 1999. Special pre-edition.
- [21] J. R. Taylor. *Linguistic Categorization*. Oxford University Press, 1995.