A Grounded Model of Reference Use: the Representation and Expression of Referring Cases by a Communicating Agent

Andrei Popescu-Belis

ISSCO, University of Geneva
Unpublished manuscript, 2001

Abstract

Numerous attempts have been made to model the phenomenon of reference to individual entities, or denotation. This paper presents a unified and grounded approach to the function of articles, inspired from French noun phrases. It is claimed that entities occupy in reality various referring cases, which an agent capable to perceive its environment may represent and express. The model attempts to simulate the emergence of function words corresponding to articles, provided the distinct expression of referring cases is relevant for the tasks required from the agents in a computer simulation.

1 Introduction and Motivation

The ability to refer is a central feature of language: it is the ability of a speaker to bring to the mind of a hearer a certain representation of an entity, about which properties are generally asserted. The typical linguistic device for reference to entities such as people or objects is the noun phrase. The study of noun phrases as *referring expressions* is a central subject in the philosophy of language as well as in computational linguistics.

B. Russell's analysis of definite and indefinite descriptions (Russell, 1905; 1919) is one of the first attempts to capture the referring function of noun phrases into a framework based on predicate logic. However, criticisms against his theory point out that in natural occurring discourse, the link between a descriptive expression and its referent is sometimes established without Russellian computation, or even against it. Donnellan (1966) proposed a distinction between the referential and attributive use of definite descriptions, and Fodor and Sag (1982) suggested a similar distinction between the referential vs. quantifier interpretation of indefinite ones.[1]

It is remarkable that despite these criticisms, most of the computational analyses of referring expressions seem to follow a Russellian, logic-based approach, as for instance the Discourse Representation Theory (Kamp & Reyle, 1993) or the Conceptual Graphs formalism (Sowa, 1984). Recent developments of the DRT attempt to grasp the functioning of various or determiners of the noun phrase, e.g. (Corblin, 1999). Another recent attempt to model the role of articles is based on an "interpretation as abduction" framework (Hobbs, 2001).

The model proposed by Appelt and Kronfeld epitomizes the success of logic-based modeling (Appelt & Kronfeld, 1987; Kronfeld, 1986; 1990), since it accommodates Donnellan's distinction into a framework based on the speech acts theory and the notion of referring acts. Yet, entities are still described by individuating sets of predicates, and it is unclear whether, despite being able to represent Donnellan's distinction, the theory is also able to predict it reliably.

Of the most interesting developments on the computational side of reference studies, algorithms for co-reference resolution (determining referring expressions that designate the same entity) have proved increasingly successful (MUC-6, 1995). The generation of noun phrases has also benefited from a logic-based description (Dale, 1992).

The theory outlined here attempts to depart from predicate logic, through an explanation of definiteness, indefiniteness, and the correlated articles within the framework of inter-agent communication. The cognitive operations underlying article production or understanding are

---

[1] This parallels a similar debate regarding the reference of proper names, and that of singular terms. The descriptive theory proposed by Frege (1892) was challenged by Kripke (1980), and more recently by causal theories, e.g. (Devitt & Sterelny, 1999).

still unclear, but the important variability in article use, as well as what is known about human categorization, motivate our search for a mechanism that is integrated into the architecture of a communicating agent.

To better explain our motivation, we will summarize first a functionalist analysis of article use (2.1) and an ineffective attempt to apply it to French while analyzing a small corpus of noun phrases (2.2). After describing the paradigm of emergent communication between computational agents (3), the theory unfolds in three parts: the referring cases occurring in reality (4), their representation in the agents' perceptual device (5), and their expression in a communication code (6). The conditions in which the distinct expression of referring cases induce a selective advantage in an agent population are finally outlined (7).

2 A Functionalist Model of Article Use and its Limits

2.1 The Bickerton-Givón Model

A simple model that accounts for the variability of article use was proposed by two functionalist linguists. The model originates in an attempt to explain the article system of Guyanese Creole and compare it to the English one, beyond their superficial similarities (Bickerton, 1981). The English-based Guyanese Creole borrowed the English definite and indefinite articles, *the* pronounced *di* and *wan* the "equivalent" of *a* (from the numeral *one*). Non-plural noun phrases with no article are much more frequent in Guyanese Creole than in English, prompting the outdated view that Creole languages are nothing but inferior versions of the superstrate language.

Bickerton's model provides an elegant explanation of the difference in article use, by introducing two features, S and P, to characterize the status with respect to reference of an entity that is referred to. The entity can be a *specific* individual or not (+S or -S) and it can be *presupposed* or not from previous discourse, i.e. known or not (+P or -P). The four possible values of the couple (S, P), namely +S+P, +S-P, -S-P and -S+P, are represented as quadrants in the S/P coordinates (Fig. 1). The respective uses of Guyanese Creole (a) and English (b) articles are represented in terms of contiguous (S, P) areas: the Creole system is simpler and somehow more

precise than the English one, which shows significant overlapping between the various uses. This representation provides an explicit account of the constraints or propensities to use various articles, which cannot be explained by simply translating them.


(a)

(b)
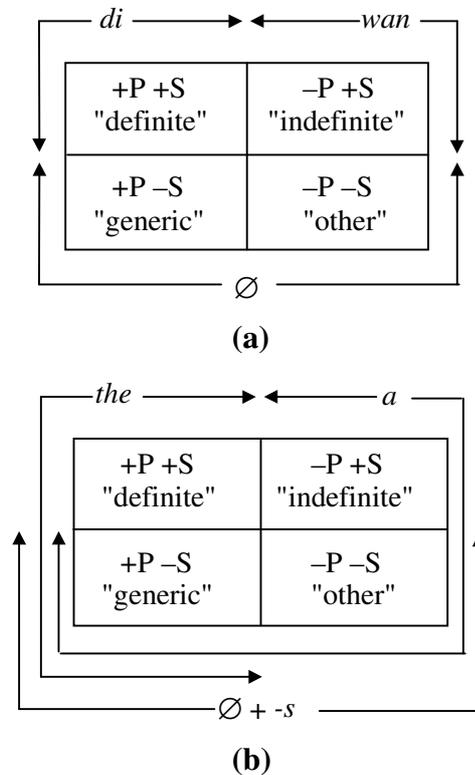
Figure 1. Article use in Guyanese Creole (a) and English (b) according to (Bickerton, 1981).

Adapting the model to more theoretical needs, T. Givón (1984:406-435) proposes other names for the four cases : "definite" (+S+P) > "referential indefinite" (+S-P) > "non referential" (-S-P) > "generic" (-S+P). He situates the cases on a helix, in which the two extreme cases are contiguous, albeit on a different level (individuals vs. categories). Givón also defends a cognitive model of reference understanding that gives a central role to the activation of the representation of referents in the working memory of a speaker. This is also a conclusion arrived at by Hopper and Thompson (1984): over a wide variety of situations and languages, articles and other nominal markers reflect the discourse salience or preeminence rather than the logical status of referents.

2.2 An Extension to French Articles

In an attempt to apply this representation to another language, we gathered and analyzed a small corpus of French non-plural noun phrases, containing about one hundred examples. The quotations of real discourse were chosen for their diversity and their capacity to illustrate non-canonical uses of definite, indefinite and no article (an excerpt is given in the Appendix). The examples were grouped first according to the articles (definite/indefinite/none). Then, an attempt was made to partition each of the three classes according to the S and P features, that is, according to the semantic, pragmatic, and discursive features of the referent.

One of the first observations is the high frequency of article omission – surprising since articles are almost compulsory in French.[2] Next, once the examples are situated in the ±S/±P coordinates (Fig. 2), the strong overlapping between the three articles becomes obvious: despite some "infrequent" uses, the definite article covers all of the four cases, the indefinite article three, the zero article two. Most important, it appears that an accurate description of the referents in the examples requires more than four cases. The proposed taxonomy is sketched in Figure 2, and fully developed in section 4.
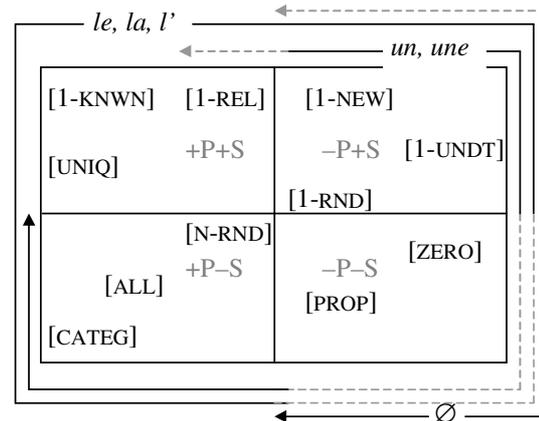
Figure 2. Extending Bickerton's model to French articles (definite: *le/la/l'*, indefinite: *un/une*). Dashed lines signal infrequent uses.

---

[2] This was far from being the case in Old French: the definite article, stemming from a Latin demonstrative adjective, was first used only for concrete nouns denoting specific individuals or objects (X[th] c.), but not when these nouns denoted non-specific individuals, or all items of a category. Then definite articles extended to this case, but were still omitted before abstract nouns (XVth c.). The contemporary use dates from the XVIIth c.; still, there are numerous cases in which the article is omitted, especially for complements of nouns or verbs (Grévisse, 1980: n. 607, 677-689). Such historical changes seem quite uneasy to grasp in a logic-based description of article use.

The strength of the functionalist representation lies in the blending of a semantic and a discursive (or pragmatic) factor: specificity (S) and presupposition (P). The four cases constitute a cognitive account of the *use* of articles, rather than of their *meaning*; they also show that articles cannot translate simply from one language to another.[3] However, in a language such as French, severe overlapping between articles tends to show that such a model must be refined to account for more than four cases. The model should also separate more clearly the objective referential properties from their cognitive representation, and better relate the latter to perception, as we shall now propose.

3 Framework: the Emergence of Communication between Agents

The multi-agent simulations of linguistic communication constitute a recent paradigm that has been applied to various problems of computational and evolutionary linguistics (Steels, 1997a), bringing new insights to problems such as the grounding of lexical items, the collective construction of meaning, the emergence and dynamics of linguistic conventions (Batali, 2000; Kirby, 1999; Steels & Kaplan, 1998; 2000). This new paradigm emphasizes the situated aspect of language use, that is, its role in an environment that the agents can perceive and modify, according to their goals and needs. Such a shift of perspective may ultimately prove essential in building a computer agent able to use language in its interaction with human users.

The phenomenon of reference is one of the keys to language grounding, since it relates language to reality through the designation of entities. But the expression of reference also inescapably involves grammar, since it makes use of common vs. proper nouns, as well as modifiers of noun phrases. The model presented below argues that the answers to current questions in reference theory as well as in agent communication lie in the encounter of the two fields.

A fundamental distinction must be made between reality, representation and expression. Quite naturally, in a multi-agent system, the

agents are able to somehow, perceive, represent and express something about the entities in their "simulated reality" using relevant symbols. Linguistic analyses are sometimes hindered by an improper distinction of these planes: for instance, it is sometimes unclear whether definiteness is a characteristic of the referent, or of its representation for the speakers, or of the description used for it (e.g., definite article).

4 Reality: the Referring Cases

Our experiment with a classification of French noun phrases (cf. 2.2 and the Appendix) prompted us to define a taxonomy of referring cases that characterize the statute of a referent (Table 1). Once again, these cases correspond to non-plural uses of nouns only.

First of all, we argue that these cases belong indeed to "reality" and not to representation. Let us adopt a (somewhat simplistic) view of "reality" in which agents evolve in a natural environment of mates, predators and preys from various species. It is of course tempting to assert that the only kind of referent are "here and now" entities, i.e. specific individuals or objects available to perception. The survival of the agents involves in fact different kinds of entities, as for instance those related to desires or needs: an agent searching for food really looks for *any entity of the category food*, and though only "here and now" entities can satisfy this need, we may say that the internal state really generated a new kind of entity. In the same way, we can argue that flight reactions can be elicited *by generic members* of predator species, which differ from reactions elicited by *known aggressive members* of one's own species. Also, an alarm call may signal the presence of *one or more predators*.

The ability to communicate increases the number of possible referring cases, adding for example *non-existent instances* of a class. It is of course difficult to defend the "reality" of such entities on metaphysical grounds; the discussion would partly parallel the famous scholastic debate on the "reality" of categories vs. that of individuals. It seems however that the case for the reality of categories finds better arguments in biology and in evolutionary cognitive science (Atran, 1990). The implementation of the category / instance distinction in an artificial environment seems feasible, for instance in a world of geometrical shapes.

---

[3] Of course the case of languages that make no use of articles (such as Japanese) must also be explained, identifying other possible devices to express reference. This is beyond our present scope.

The list of referring cases (henceforth, RCs) that we identify in naturally occurring noun phrases is given in Table 1, and their position with respect to the ±S/±P characterization appears in Figure 2. The order in which these cases are listed parallels Givón's hierarchy, with the two extremes corresponding to a "non referential" use, on different planes.

While our RCs attempt to cover all possible uses of non-plural noun phrases, their limits are fuzzy and they sometimes overlap. In other words, when classifying the use of a noun phrase in context, it is highly improbable that it fits none of the RCs, but there may be some hesitation between two neighboring RCs.

Table 1. Names / definitions of referring cases.

[PROP]	the NP evokes a *property* of the category rather than an instance
[ZERO]	the NP is used to assert that something holds *for no instance* of the category
[UNIQ]	the NP evokes the *unique* and known instance of the category
[1-KNWN]	one instance, already introduced in the discourse and thus *known*
[1-REL]	an instance that is *related* to an already introduced instance (NP complement, relative clause)
[1-NEW]	a *new* instance
[1-UNDT]	a *new* instance that is still *undetermined*
[1-RND]	*one and only one random* instance of the category
[N-RND]	*one or more random* instances
[ALL]	*any* instance or *all* instances
[CATEG]	the *category* itself, as an instance of a higher-order category

One may wonder whether the definitions of the RCs would not receive a better formulation in predicate logic. This is probably true for at least some of the RCs ([UNIQ], [ZERO] [1-UNDT]), but clearly not for all of them ([PROP], [N-RND]). A logic-based approach of reference seems tractable for rational agents (Appelt & Kronfeld, 1987; Kronfeld, 1986; 1990), but not for the perceptive agents used in simulations of language evolution. Our approach departs at this point from symbolic representations, as the RCs will receive below proper definitions in terms of perceptual representations, which are then related to expression; the definitions provided in Table 1 pertain thus more to explanations.

5 The Representation of Entities

The central idea of this section is that agents possess a perceptual device maintaining different representations for the features pertaining to individuals and those pertaining to categories (5.1). We hypothesize that RCs are represented in terms of the differential activation of these two representations (5.2). Further hypotheses are necessary to account for the speaker / hearer distinction and the new referent / old referent distinction (5.3).

5.1 Categorization and Memory

The categorization mechanism implemented in the *Darwin II* model (Edelman & Reeke, 1982) is based on two parallel pathways: the (I) pathway detects and represents the properties of instances, while the (C) pathway detects and represents the category of each instance. Since only one entity is perceived at a time, there is no need for a binding or integration mechanism between the various features in the two pathways.

To get closer to a model for a perceptive agent, we will use for our proposal value intervals in the perceptual channels as features, as for instance in (Steels, 1997b). Let us suppose that the perceptual input is decomposed into a series of "channels" or scalar variables, with values between a minimal and a maximal value. Each channel is monitored by two first-order detectors, one fine-grained, for the properties pertaining to individuals, and the other coarse-grained, for the properties pertaining to categories. The first channel divides the interval of possible values into small sub-intervals, each representing a possible value of the respective feature, while the latter uses larger sub-intervals.

A category detector is probably present in all vertebrates, while an instance detector appears only in more evolved species. The experiments described in (Steels, 1997b) focus on the emergence of an instance detector, that is the emer-

gence of conventions for the relevant intervals and the associated lexical items. Incidentally, the lack of a category detector has been imagined by J.L. Borges in *Funes el Memorioso* (from *Ficciones*).

Second order detectors must integrate the conjunction of various feature values in order to

build a representation of the entities. Our proposal here is that the features are also characterized by an activation or confidence value, a result of an uncertainty (noise) distribution on
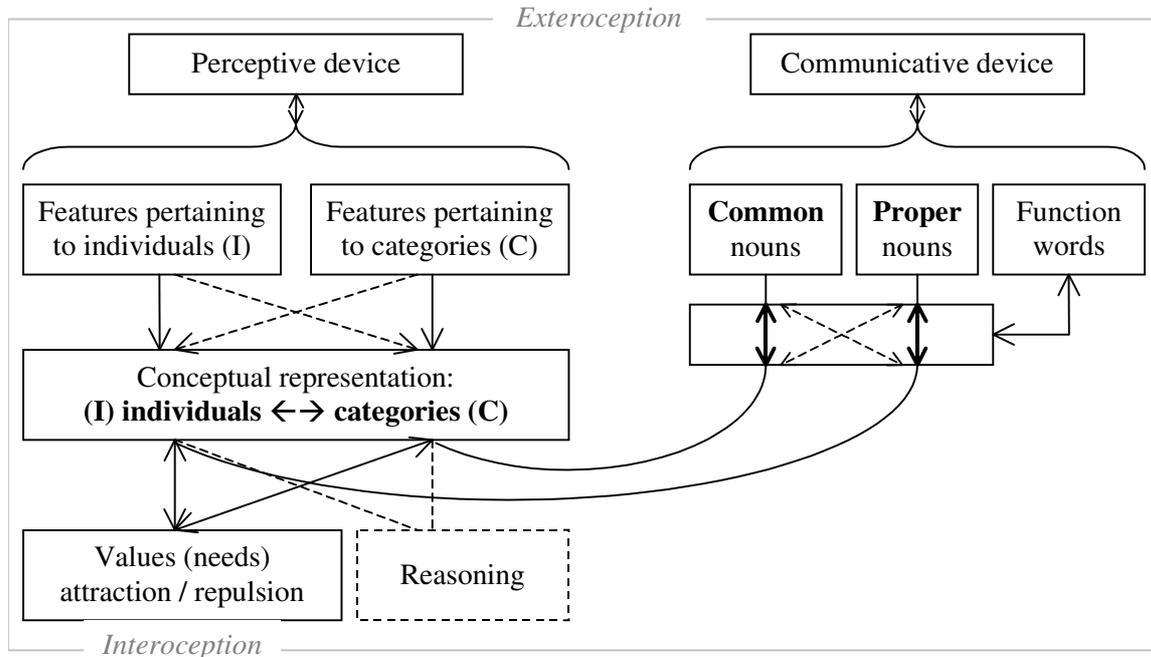
**Figure 3**. Structure of an agent that is able to represent and express referring cases.

the input channels. It is essential to note that two second order representations may possess the same set of individual (I) and categorial (C) features, and still differ in the relative activation of these features. The second order representations correspond thus to "concepts" of categories, and respectively, individuals. The various detectors are schematized in Figure 3, together with the other elements of the proposed architecture.

5.2 Representation of Referring Cases

We believe that the conjunction of (I) and (C) features enables an agent to represent the RCs of the perceived entities. The parameters that govern this representation are the respective total activations of (I) and (C) features; "total" means the sum of channel activations. The (I)+(C) total may be interpreted as the global salience, in the working memory, of the respective entity.

Each of the RCs corresponds thus to the region of the possible activation values in the (I) / (C) plane. There are no clear-cut boundaries between these regions: rather, the various RCs

situated on the graph in Figure 4 indicate only the center of regions whose borders interpenetrate. The centers of regions corresponding to the ±S/±P classification are also shown.
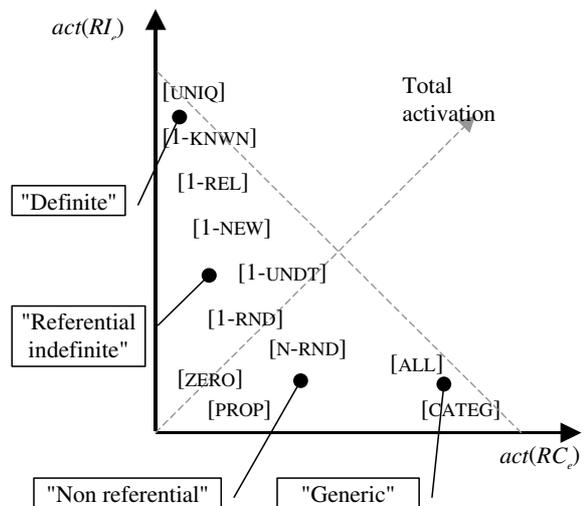
Figure 4. The relation between referring cases and (I)/(C) total activations. The four cases of the ±S/±P classification are also represented.

The description provided here is the *operational* definition of RCs, as opposed to the explanations in Table 1. For instance, the representations of a well-known unique instance, or one related to a known one, or of a new instance, correspond to a high (I) activation ([UNIQ] > [1-KNWN] > [1-REL] > [1-NEW]). On the other hand, the representation of a category, or of "any" instance, or of random instances, has a high (C) activation ([CATEG] > [ALL] > [N-RND]). Finally, [ZERO] and [PROP] cases have low activation on both scales.

5.3 Two Distinctions

As defined here, the representation of RCs seems to be invariant with respect to the speaker and the hearer – no such distinction has been made yet. How can the model accommodate the informational asymmetry between speaker and hearer without modifying the representation of RCs? We suggest here that the working memory storing the representations of referents (i.e. the distribution of feature activations) should be viewed as a *memory of the communicative interaction* between agents, that is, a memory of what has been said and thus became common knowledge. If reference was solved unambiguously by the speaker and the hearer – which is supposedly the case if no anomaly or reference negotiation occurred – then the memories contain similar representations.

There is a stage, of course, when the two memories are not synchronized: in the interval between the thought of a new referent occurring to the speaker and the decoding of the corresponding referring expression by the hearer. The first event is commented in section 7, while the understanding of the second relies on the analysis of reference resolution by the hearer (how to decide whether a referent is new or old) that is beyond our present scope.

6 Expression of Referring Cases

The last step of this model is the link between the RCs and the vocabulary of the agents (more about the emergence of such a vocabulary in section 7). The basic idea underlying such a vo-

cabulary is that the agents are able to emit and receive a given repertoire of words, which are divided among three classes. Proper nouns are associated to concepts of individuals and common nouns are associated to concepts of categories (Fig. 3).

Our main claim at this point is that articles are devices (or *function words*) that enable a *content word* to signal various referring cases for a given concept. For instance, an article will allow a word associated to a categorial concept (a common noun) to evoke an individual member of the category, and even express whether this individual is a determined one or a random instance of the category. Conversely, a word associated to an individual (a proper noun) could also be modified by articles to evoke a category, or another entity from such a category (this use is less frequent in natural language).

Focusing now on the expression of the various RCs using non-plural common nouns, we hypothesize that the various regions of the (C) / (I) plan correspond to propensities in the use of articles, as shown in Figure 5 for French articles. The zone with the highest total activation corresponds to a dominant use of the definite article, the next zone to the indefinite article, and the lowest activation zone to article omission. Ideally, each region corresponding to an RC should possess its own marker (function word), but only three markers exist in French, which explains why the regions we propose for each of them encompass several RCs. Once again, the borders of these regions are fuzzy, the diagram being only an idealization. Such a diagram is also useful to explain differences in article use, e.g. between French and English, by differences in the areas covered by the articles.

$act(RI_e)$

le, la, l'

un, une

∅

$act(RC_e)$
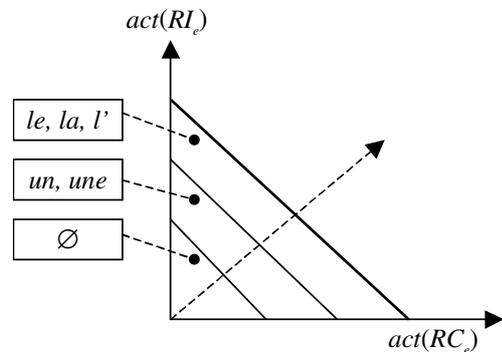
Figure 5. The propensity to use French definite / indefinite / zero article with common names for the various referring cases.

7 Towards the Use of the Ability to Communicate Referring Cases

7.1 Connection to Values

In order to witness the emergence of function words to express referring cases, agents should have some interest to do so. The first step in this direction is the implementation of values or needs in an agent. The external entities should be associated to internal states, pleasure or satisfaction triggering attraction, and fear or pain triggering repulsion. This would constitute the basis for learning associations between individuals or categories and various internal states, using for instance associative learning between the perceptual and interoceptive detectors.

The automatic activation of certain internal needs would evoke in turn the associated representations, providing a reason to communicate about more than "here and now" entities, that is, about entities in various RCs, not always available to perception. For instance, the activation of an (I) concept by an internal need would trigger a random activation of (I)-like features derived from the corresponding category, while the activation of a (C) concept would trigger mostly an activation of the (C) features.

7.2 Experimental Prospects

The motivation for the emergence of function words to express RCs in agent communication would be based on two factors. First, a constraint on the (I) conceptual repertoire would limit the number of possible proper nouns, thus increasing the pressure for common noun use.

Second, two forms of learning should be possible: one by experience (experimenting pleasure, pain, etc.) and one by communication. Also, among the entities present in the environment the agents should find attractive / repulsive species as well as individuals. This would induce a selective advantage on the ability to correctly express the RCs. Even if only a few RCs are modeled at the beginning, such an experiment would bring the first piece of evidence to the emergence of a grammatical device.

References

Appelt, D., & Kronfeld, A. (1987). *A Computational Model of Referring*. IJCAI '87, Milan.

Atran, S. (1990). *The cognitive foundations of natural history*. Cambridge, UK: Cambridge University Press.

Batali, J. (2000). The Negotiation and Acquisition of Recursive Grammars as a Result of Competition Among Exemplars. In T. Briscoe (Ed.), *Linguistic Evolution Through Language Acquisition*. Cambridge, UK: Cambridge University Press.

Bickerton, D. (1981). *Roots of Language*. Ann Arbor, Michigan: Karoma Press.

Corblin, F. (1999). Mentional references and familiarity break. In M. Coene (Ed.), *Hommages à Liliane Tasmowski*, (pp. 535-544). Padova: Unipress.

Dale, R. (1992). *Generating Referring Expressions: Building Descriptions in a Domain of Objects and Processes*. Cambridge, MA: The MIT Press.

Devitt, M., & Sterelny, K. (1999). *Language and Reality: an Introduction to the Philosophy of Language*. Cambridge, MA: The MIT Press.

Donnellan, K. S. (1966). Reference and Definite Descriptions. *Philosophical Review, 75*, 281-304.

Edelman, G. M., & Reeke, G. N. (1982). Selective networks capable of representative transformations, limited generalizations and associative memory. *PNAS,* 79, 2091-2095.

Fodor, J. D., & Sag, I. A. (1982). Referential and Quantificational Indefinites. *Linguistics and Philosophy, 5*, 355-398.

Frege, G. (1892/1956). On sense and reference. In P. Geach & M. Black (Eds.), *Translations from the Philosophical Writings of Gottlob Frege*. Oxford: Blackwell.

Givón, T. (1984). *Syntax: a functional-typological introduction*. Amsterdam: John Benjamins.

Grévisse, M. (1980). *Le Bon Usage*. (11th ed.). Paris-Gembloux: Duculot.

Hobbs, J. (2001). *"the", "a" and ""*. International Workshop on Reference and Coherence in Discourse, Utrecht.

Hopper, P. J., & Thompson, S. A. (1984). The discourse basis for lexical categories in universal grammar. *Language,* 60(4), 703-752.

Kamp, H., & Reyle, U. (1993). *From Discourse to Logic*. Dordrecht: Kluwer.

Kirby, S. (1999). *Function, selection, and innateness: the emergence of language universals*. Oxford, UK: Oxford University Press.

Kripke, S. A. (1980). *Naming and Necessity*. Cambridge, MA.: Harvard University Press.

Kronfeld, A. (1986,). *Donnellan's Distinction and a Computational Model of Reference.* 24th Meeting of the ACL, New York, NY.

Kronfeld, A. (1990). *Reference and computation: an essay in applied philosophy of language.* Cambridge, UK: Cambridge University Press.

MUC-6. (1995). *Proceedings of the 6th Message Understanding Conference.* San Francisco, CA: Morgan Kaufman.

Russell, B. (1905). On Denoting. *Mind, 14,* 479-493.

Russell, B. (1919). Descriptions. In *Introduction to Mathematical Philosophy.* London: George Allen and Unwin.

Sowa, J. (1984). *Conceptual structures: information processing in mind and machine.* Reading, MA: Addison-Wesley.

Steels, L. (1997a). The Synthetic Modeling of Language Origins. *Evolution of Communication Journal*, 1(1), 1-34.

Steels, L. (1997b). *Constructing and Sharing Perceptual Distinctions.* European Conference on Machine Learning.

Steels, L., & Kaplan, F. (1998). *Spontaneous Lexicon Change.* Coling-ACL '98.

Steels, L., & Kaplan, F. (2000). Bootstrapping Grounded Word Semantics. In T. Briscoe (Ed.), *Linguistic evolution through language acquisition.* Cambridge, UK: Cambridge University Press.

Appendix

These excerpts of a corpus of French noun phrases are sorted according to their referring cases. For each case, the possibility to use a definite article, an indefinite one, or no article, is exemplified. For lack of space the examples could not be translated into English.

[PROP] – Le chien est *l'ami* de l'homme./ Paul est *un docteur* qui connaît son métier./ Il fait un temps de *chien*. Paul est *docteur*.

[ZERO] – Jean n'a pas rencontré *le lion* en Afrique./ Il n'y avait plus *un chien* dehors./ Jean n'a pas vu de *lion* en Afrique.

[UNIQ] – *Le soleil* se couche.

[1-KNWN] – J'ai acheté un chien et un chat. *Le chien* est très amical.

[1-REL] – Pierre s'est cassé *la jambe*./ Pierre s'est cassé *une jambe*./ Le procès est fini. *Accusateur* et *accusé* ont été critiqués par le juge.

[1-NEW] – Je suis allé hier voir *le docteur* : tout va bien. Après *la bataille*, *le roi* parla ainsi *au poète*…/ J'ai acheté *un chien* et *un chat*./ Nous nous sommes mis à *table*.

[1-UNDT] – Paul lit *le journal*./ Jean cherche *une femme*. / Que serions-nous sans *femme* ?

[1-RND] – *Le médecin* peut réussir là où *le prêtre* échoue./ Il a été traité comme *un chien*./ La mère mit l'enfant dans une peau de *renard*.

[N-RND] – Jean est parti chasser *le renard*./ Il y a une odeur de *chien* dans cette pièce.

[ALL] – *Le lion* est carnivore. / *Une mère* peut-elle ne pas aimer son enfant ?

[CATEG] – *Le chien* fait partie de la classe des vertébrés.