

Semiotic Dynamics for Embodied Agents

Luc Steels, *Vrije Universiteit Brussel and Sony Computer Science Laboratory Paris*

Artificial intelligence explores the many different aspects of intelligence like a meandering stream carving out rivers, lakes, and deltas in an endless magnificent landscape. Each time new vistas on intelligence open up, we build new technologies to explore them and find new types of applications.

Through empirical study, computational modeling, and theoretical study, we can learn more about a new branch of science—semiotic dynamics—that offers insight into the nature of human communication.

What recent developments in AI have fired the imagination in this first decade of the 21st century? I argue that it's the study of *semiotic dynamics*: the processes whereby groups of people or artificial agents collectively invent and negotiate shared semiotic systems, which they use for communication or information organization. Tagging systems (such as Flickr, CiteULike, del.icio.us, or connotea) offer examples of human semiotic dynamics at work, aided by technologies such as the Internet but also by a new sense of collective action in an increasingly connected world. Human semiotic dynamics also result in natural languages, which have played a role similar to collaborative tagging in coordinating and shaping human activity since our species' birth.

Semiotic dynamics builds on many earlier AI developments: the insights into and technologies of semantic networks and knowledge representation from the seventies, the ideas on embodiment and grounding from the late eighties, and the perspective of multiagent systems from the nineties. But all these aspects join together into a new vision on intelligence, with the social, collective dynamics of representation-making at the center. These new AI developments don't stand in isolation. They resonate with recent developments in linguistics and psychology, where the research focus is rapidly shifting from looking only at the (static) competence of an "idealized speaker" toward the notion that language is a complex adaptive system, continuously changing and adapting to accommodate its users' needs in social interaction. They rest also on new developments in the mathematical study of networks,

which have triggered rapid advances in understanding the behavior of numerous complex systems, from stock markets to ecologies to the Internet. Here, I briefly illustrate the current study of semiotic dynamics, the resulting technologies, and the field's impact on current and future intelligent systems applications.

Language as a complex adaptive system

In the late eighties and early nineties, psycholinguists started looking in detail at situated human dialogue and came to a startling conclusion.¹ Dialogue partners quickly and continuously adapt to each other at all levels of language: phonetic, lexical, syntactic, grammatical, conceptual, and pragmatic. Of course, they come to the conversation with partly established systems, but these aren't fixed systems; they're adapted, expanded, and negotiated to serve the ongoing dialogue—often based on repairing miscommunications²—to optimize communicative success and minimize cognitive effort. Such change might involve

- inventing new words or stretching usage,
- adopting a grammatical construction for a slightly new purpose,
- introducing new concepts or shifts in how concepts are perceptually grounded in sensorimotor experience,
- aligning dialogue partners' speech sounds and intonation patterns, and
- coordinating turn-taking behavior and gestures.

Graphical semiotic systems that use drawings or other visual media to communicate meaning undergo similar fast changes.³

These findings startle because the standard view had been that language was basically static, barring minor variations due to performance. Most people assumed that we acquire language during childhood and that it remains more or less constant throughout life. Some linguists argue that the system is largely universal (particularly the sound, grammatical, and conceptual systems) and that it's genetically coded and transmitted. So, they believe that a language learner need only fill in the lexicon and fine-tune the syntactic parameters.⁴ Others believe that you can learn a language system without any social interaction by induction from the kind of data typically available to a child, and, once learned, you use it without significant modification throughout life.⁵

This traditional *iterated transmission* view has informed decades of research in computational linguistics and the machine learning of language. Researchers have tried to find the statistical learning processes that could extract “the language system” from a corpus of data.⁶ Others have tried to identify “the” set of innate, universal primitive concepts and relations⁷ or map out an encyclopedic view of the universal foundations of commonsense knowledge.⁸ The iterated-transmission view has also given shape to the design and implementation of intelligent user interfaces (whether natural language based or not). So, today's interfaces are typically nonadaptive in that they assume fixed, familiar rules for interaction that those interacting with the interface can learn.

But apparently, a static universal language system isn't the norm for human languages, or human semiotic systems in general. Natural semiotic systems are forever emergent, only partially shared, and usage based. With this in mind, can we devise artificial systems with the same type of semiotic adaptability as humans that could then participate in the collective construction and propagation of languages—themselves complex and adaptive living systems?

The naming game

The earliest attempts to build agent models that exhibit the highly flexible semiotic dynamics seen in human groups date from the mid-1990s. These attempts first occurred in an artificial-life context, partly to understand and model the origins of languages grounded

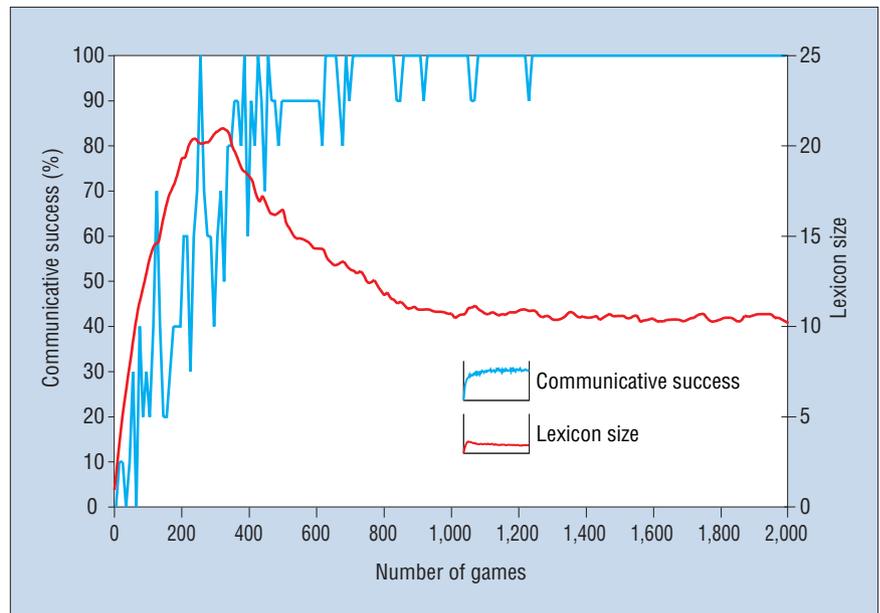


Figure 1. Semiotic dynamics of the naming game for 10 agents and 10 objects. After only 500 games (25 per agent), communicative success already exceeds 90 percent.

in the sensorimotor experiences of situated embodied agents.⁹ One such model, the *naming game*, plays a similar role in the study of semiotic dynamics as the prisoner's dilemma game does in sociology and economics.

The naming game is played by a population of agents—having the same sort of cognitive mechanisms but each with its own local memory. There's no telepathy, central control, or prior vocabulary of names. In each game, two random players are chosen from the population. The speaker picks an object from the current context, searches his memory for a name for this object, and tells the hearer. The hearer then points to the object that she associates with this name. The game is a success if the hearer correctly identifies the object that the speaker originally chose. Otherwise, it's a failure and the speaker points to the object he intended.

Many strategies exist for this sort of game. For example, agents could use a neural network-like associative memory for storing weighted associations between objects and names (ignoring the issues of object recognition and speech production or perception) and increase or decrease the weights on the basis of success or failure in the game. However, any strategy chosen must have two basic ingredients: alignment and innovation.

After every game, the speaker and hearer should align or adjust their associative memories to make their vocabularies more similar. For example, suppose they used weighted

networks. If the game is a success, they should increase the weight of the used association so that it's used even more in the future and decrease the weight of competing associations (other names for the same object). If the game is a failure, the agents know that they don't share each other's associations, so they should diminish their respective weights.

In addition to alignment, we need innovation—otherwise, the system could never move forward. Innovation here means that a speaker who doesn't have a name for an object could invent one and add it to her memory. If the hearer doesn't know a name, he can adopt it by storing it in his memory with some initial weight.

A strategy based on these two ingredients quickly leads a group of agents to a set of shared naming conventions without global control or prior design (see figure 1). Researchers have also studied this game using a more classical iterated-transmission approach, both genetically¹⁰ and culturally,¹¹ but the model I discuss here relies entirely on self-organization due to the positive-feedback loop between use and success. Agents don't need to go through many generations before reaching a consensus, so convergence appears surprisingly fast. Moreover, the system remains adaptive: When new objects appear, agents invent names for them, which then spread throughout the population. When new agents enter the population, they might

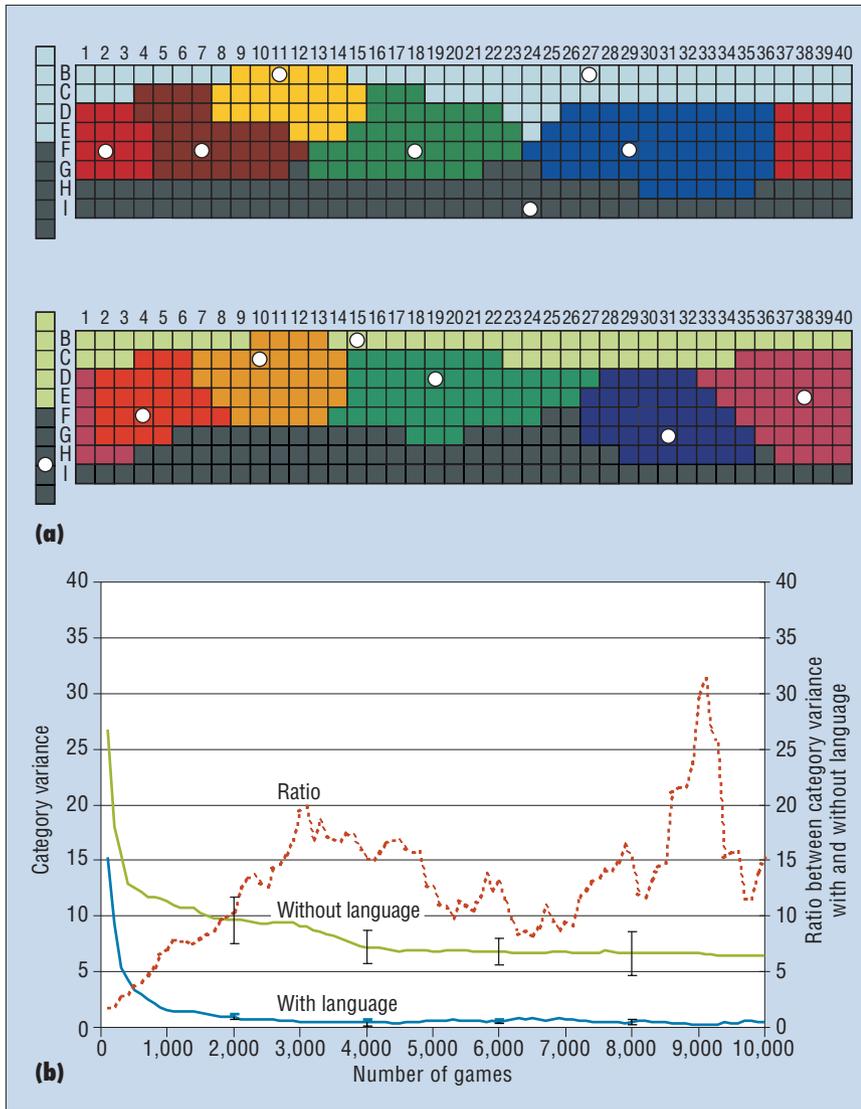


Figure 2. Agents develop a repertoire of color prototypes to distinguish color chips. The (a) repertoire of two different agents and the (b) category variance with and without category formation coupled to language: when agents name colors and use them in this guessing game, the color prototypes become more similar.

occasionally invent names, but the higher frequency of existing conventions lets the names remain stable and permits vertical transmission from one generation to the next. When agents leave, enough agents still remain to act as a collective memory.

Since the early research on the naming game a decade ago, researchers have defined and studied much more complex games through computer simulations. The *guessing game* goes a step up in complexity; it no longer associates a name with an individual object but with a perceptually grounded category that can pick out an object from the context.¹² Examples include red versus orange or

small versus big. If all agents in the population start with a fixed set of categories, we have the same dynamic as the naming game. But if the agents have open-ended ontologies, a dual dynamic appears. Categories of individuals agents become more similar and the terms for naming these categories become increasingly shared. The two dynamic systems—the evolving ontologies and evolving lexicons—are coupled to each other.

The game encourages successful solutions, both at the conceptual and lexical levels, and suppresses competing solutions or choices that don't work. Many computer simulations and experiments have shown that

this dynamic lets a population self-organize a shared lexicon as well as the underlying perceptually grounded ontology. The principles of alignment and invention are thus carried to the conceptual level. (Figure 2 shows an experiment in the conceptual alignment of color categories.¹³) Agents impose (or invent) categorical boundaries but shift them if they can't distinguish the topic from other objects in the context or if particular boundaries aren't successful in the guessing game.

Similar multiagent modeling has shown that alignment and invention can explain how a population can self-organize a shared repertoire of speech sounds¹⁴ and an increasing number of experiments apply the same approach to study the emergence of grammar.¹⁵ You can find more on this topic in recent review articles and book collections.^{16,17}

Complex systems theory

Although we know from computer simulations that a population of embodied agents can self-organize a semiotic system, we've only recently understood why this was or how to predict these systems' properties, such as convergence time, scaling laws, limits in population or object in- and outflow, resistance to noise in transmission or pointing, and so on. It turns out that we needed a set of mathematical tools developed recently as part of advances in graph theory triggered by the study of complex networks and their application to a wide range of complex dynamical systems.

Andrea Baronchelli, Vittorio Loreto, and their colleagues recently discovered how to apply these techniques to semiotic dynamics.¹⁸ They studied a simplified naming-game strategy (a more discrete version of the weighted updating discussed earlier) and population scaling. They discovered that the convergence times (t_{conv} , which generally follow an S-shaped curve typical in studies of language change) scale following a power law (see figure 3). Building on this, we might ask how we can explain this power law's exponent and use this to understand the naming game's collective dynamics. Here, we can learn from complex-systems science. We can aggregate the various quantities involved (for example, the number of words known by the population's agents for an object) and develop master equations regulating them. Adopting the mean-field assumption common in statistical mechanics, we can then predict the system's general behavior, including the power law's exponents.¹⁸

Moreover, by focusing on a single meaning, we can imagine that all agents that have a word in common are linked, so that a fully connected clique exists for every word. The agents' lexicons of the agents reach total *coherence* (obtained when all agents use the same word for the same meaning) when the population's agents form a fully connected graph without multiple links. By viewing distributed language systems as networks, we can understand them in terms of processes for network growth and decay, such as the rich-get-richer dynamics of small-world networks.¹⁹

Applying complex-systems techniques to other language games and other aspects of language, such as phonetic convergence, is an active field with much to discover. This theoretical research has already helped to design more efficient agent behaviors, to bring to light aspects overlooked in earlier simulations or to clearly establish the field's foundations. For example, we now know that network topology is critical in determining whether and how fast populations will converge toward the same set of conventions.²⁰ Intuitively, hubs (densely connected nodes) have a greater influence than relatively isolated nodes. It's perhaps less intuitive that hubs have benefits—each agent needs significantly less memory to play the game—and drawbacks—convergence times become slower as the group reaches total convergence. We can expect to see many more results of this type in the near future.

Given all these computational and theoretical insights, what can we do with them in creating intelligent systems? We'll look at two areas that have explored this question: emergent communication systems for physical robots and semiotic dynamics on the Web.

Communication systems for robotic agents

From the beginning, semiotic dynamics research was motivated by understanding how distributed autonomous robots could self-organize communication systems using their own sensorimotor experiences of the world (not necessarily shared by humans). It's particularly appropriate for robots to have an open-ended, adaptive communication system when they must deal with unknown, open-ended environments. For example, if we send a robot colony to Mars, we can't control them remotely. So, they need some sort of communication system—particularly if they separate to explore the planet or divide up labor. When

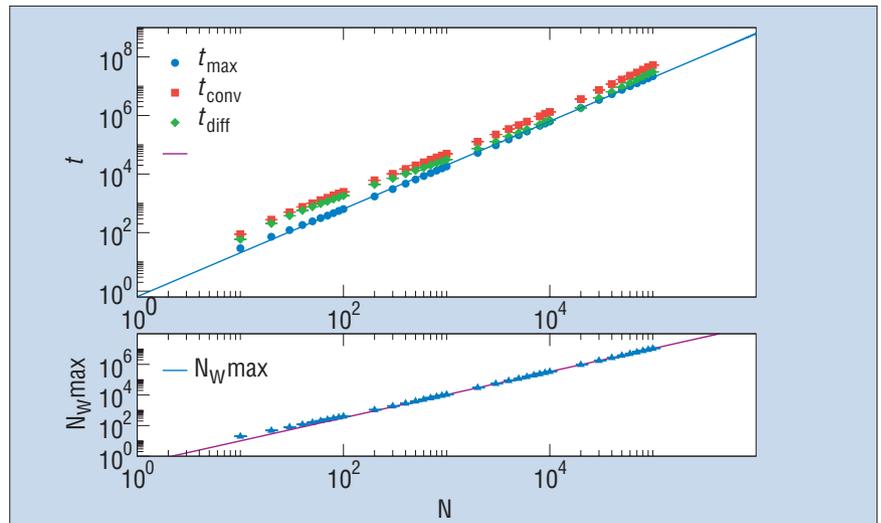


Figure 3. A log-log plot showing the naming game's power law behavior. The x-axis shows population size; the y-axis show t_{max} , the time when the number of words is at a maximum, and t_{conv} , the time when the population reaches convergence. Both show power law behavior with exponent 1.5.

we can't predict all situations the robots might want to discuss, the communication system must be open ended, particularly if we don't share the same sensory-motor Umwelt.

The language games I've discussed weren't only developed in simulation but were also tested with visual and motor stimuli in real-world environments, with some remarkable results. Despite the noise and uncertainties, the agents' alignment and invention behaviors still led to the same self-organized emergence of semiotic systems seen in computer simulations. This showed that we can find behaviors both fast and robust enough for practical application. In one test, the Talking Heads Experiment, agents used two pan-tilt cameras to play guessing games about geometric figures on a whiteboard in front of them.²¹ Both a shared ontology (with perceptually grounded categories such as red, left, small, and so on) and a shared vocabulary emerged without central control.

More recently, a perspective-reversal experiment (see figure 4) nicely illustrated the state of the art.²² It involves freely moving Sony AIBO robots that describe to each other a recent event involving a ball. However, the robots usually look at the scene from different points of view, so they never have a shared perception of the world, requiring a reference-frame alignment—the speaker aligning to the hearer's perspective or vice versa. For example, one robot could say, "There is a ball rolling from your left toward me." This lan-

guage game would be a success if the robot-speaker's description fits with the robot-hearer's perception of the situation. The AIBOs were members of a larger population and didn't have a built-in set of conventions or ontology. Moreover, the experiment aimed to examine whether autonomous robots could use egocentric-perspective transformations to reconstruct what the world looks like from another agent's viewpoint and thus align reference frames in a scene's conceptualization.

Figure 4b illustrates the experiment's success: communicative success is above 80 percent, and 50 words are consistently used in their lexicons. Not only do perceptually grounded categories (such as close by, from left to right, move fast, and so on) emerge and get consistently lexicalized, we also see emerging perspective markers combined with categories—for example, "rolling from my left to your right." The experiment thus explains perspective marking's prevalence in human languages. Remarkably, the semiotic dynamics based on alignment and invention (both at the conceptual and lexical levels) self-organized a semiotic system in a group of robots, despite the noise and uncertainties of real-world vision and physical behavior.

These experiments show that artificial systems can handle the symbol-grounding problem—although philosophers have argued this was impossible—when we take a social view on symbol emergence.

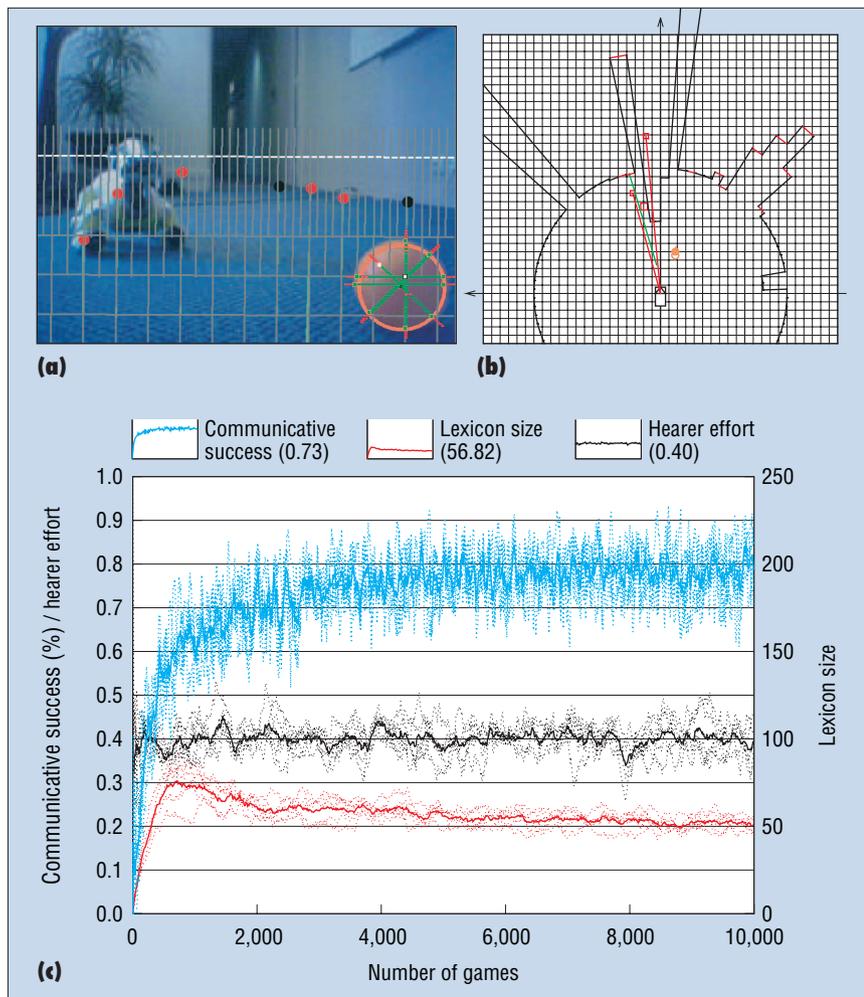


Figure 4. A reference-frame alignment experiment using autonomous AIBO robots: (a) one robot’s view of the other robot, (b) the ball-and-world model showing obstacles and ball trajectory, and (c) the semiotic dynamics of an experiment involving five robots playing 10,000 language games. Hearer effort measures the cognitive effort, which agents try to minimize.

From collaborative tagging to emergent semantics

Collaborative tagging is another exciting application of semiotic dynamics that has recently taken off. Social-tagging sites such as Flickr (see figure 5a) or CiteULike let users upload media objects (pictures, Web pages, texts, videos, music files, and so on) and tag them with freely chosen keywords. Users can also organize themselves in social networks with objects local to friends (as in peer-to-peer networks) or entirely public. Although users’ tags are fully distributed and driven by their or their closest peers’ needs, we see a remarkably stable “folksonomy” emerge (see figure 5b²³). This resembles the language games’ dynamics. A winner-take-all phenomenon typically

occurs with pure synonyms (such as New York versus NY, NYC, or New York City), but new tags can occasionally overtake others owing to important changes in the world or new fashions. We can study these systems’ semiotic dynamics using the same tools as for the language games. For example, we can explain the appearance of various power laws and distributions using statistical models such as Polya processes.

Current collaborative-tagging sites use state-of-the-art Web technologies, including well-designed user interfaces and distributed databases. They don’t incorporate significant intelligent technologies so far, but that’s only because they’re new. On the basis of earlier developments in AI, we can expect

possible evolution in three directions:

Co-occurrence. Much of the earlier work in AI has shown that we can mine the human symbol system’s statistical structure—for example, by computing the co-occurrence of words in texts (as in semantic latent analysis). These techniques have potentially high relevance for providing tagging systems with greater flexibility and are, indeed, progressively being incorporated. We can compute the clustering of tags based on co-occurrence and present users *tag clusters* (also known as tag clouds) as an additional basis for browsing.

Grounded semantics. At this point, tagging systems have no grounded semantics except for a user’s declaration that an item belongs to a set associated with a tag. Grounded semantics would mean associating classifiers with tags (for example, you could associate a classifier that can detect whether a picture is black and white with the tag “black-and-white”). Here, the vast amount of earlier pattern recognition and machine learning research becomes relevant, because, in principle, it’s possible to learn the classifier on the basis of examples and counterexamples of images or music files associated with particular tags. A grounded semantics could help deal with synonyms (“black-and-white” is sometimes also called “black-white” or “bw”) or even crosslinguistic use of tags (such as “noir-et-blanc” to mean black-and-white pictures). It could also help make valuable suggestions for users so they don’t have to tag every item explicitly and weed out tags that seem totally inappropriate (for example, an image without any red tagged “red”).

Inference. Another relevant AI area relates to knowledge representation and inference. The Semantic Web has already introduced ideas from logical inference into Web technologies, specifically by designing tools for defining ontologies and ontologies to use in information organization and access. Tagging sites more fully exploits the bottom-up emergent properties characteristic of human semiotic dynamics, but undoubtedly we can find ways to use certain inference types to help browse through the space of tags.

In this new field of semiotic dynamics, much is undiscovered—particularly for understanding higher-order conceptualiza-

tions and verbalizations involving emergent grammar. Nevertheless, huge potential for application already exists. We can consider new ways to support the highly dynamic forms of human semiotic dynamics, as shown in collaborative-tagging sites or natural language communication. We can start to think seriously about adaptive interfaces, which aren't based on static interaction protocols but dynamically negotiated in cooperation with users. Finally, we can envision robotic applications with the flexibility to deploy more powerful communication systems in open-ended, unknown environments. ■

Acknowledgments

The ECAgents project—funded by the Future and Emerging Technologies program (IST-FET) of the European Commission under EU RD contract IST-1940—partly supported this research, conducted at the Sony Computer Science Laboratory Paris. The information provided is the author's sole responsibility and doesn't reflect the Commission's opinions. The Commission is not responsible for any use that may be made of data appearing in this article.

References

1. S. Garrod and A. Anderson, "Saying What You Mean in Dialogue: A Study in Conceptual and Semantic Coordination," *Cognition*, vol. 27, no. 2, 1987, pp. 181–218.
2. H.H. Clark and S.A. Brennan, "Grounding in Communication," *Perspectives on Socially Shared Cognition*, L.B. Resnick, J.M. Levine, and S.D. Teasley, eds., APA Books, 1991.
3. P. Healey et al., "Graphical Representation in Graphical Dialogue," *Int'l. J. Human-Computer Studies*, vol. 57, no. 4, 2002, pp. 375–395.
4. N. Chomsky, *Lectures on Government and Binding*, Foris Publications, 1981.
5. J. Elman, "Distributed Representations, Simple Recurrent Networks, and Grammatical Structure," *Machine Learning*, vol. 7, nos. 2–3, 1991, pp. 195–226.
6. E. Charniak, *Statistical Language Learning*, MIT Press, 1993.
7. R.C. Schank, *Conceptual Information Processing*, Elsevier, 1975.
8. D.B. Lenat, "Cyc: A Large-Scale Investment in Knowledge Infrastructure," *Comm. ACM*, vol. 38, no. 11, 1995, pp. 33–38.
9. L. Steels, "The Synthetic Modeling of Lan-

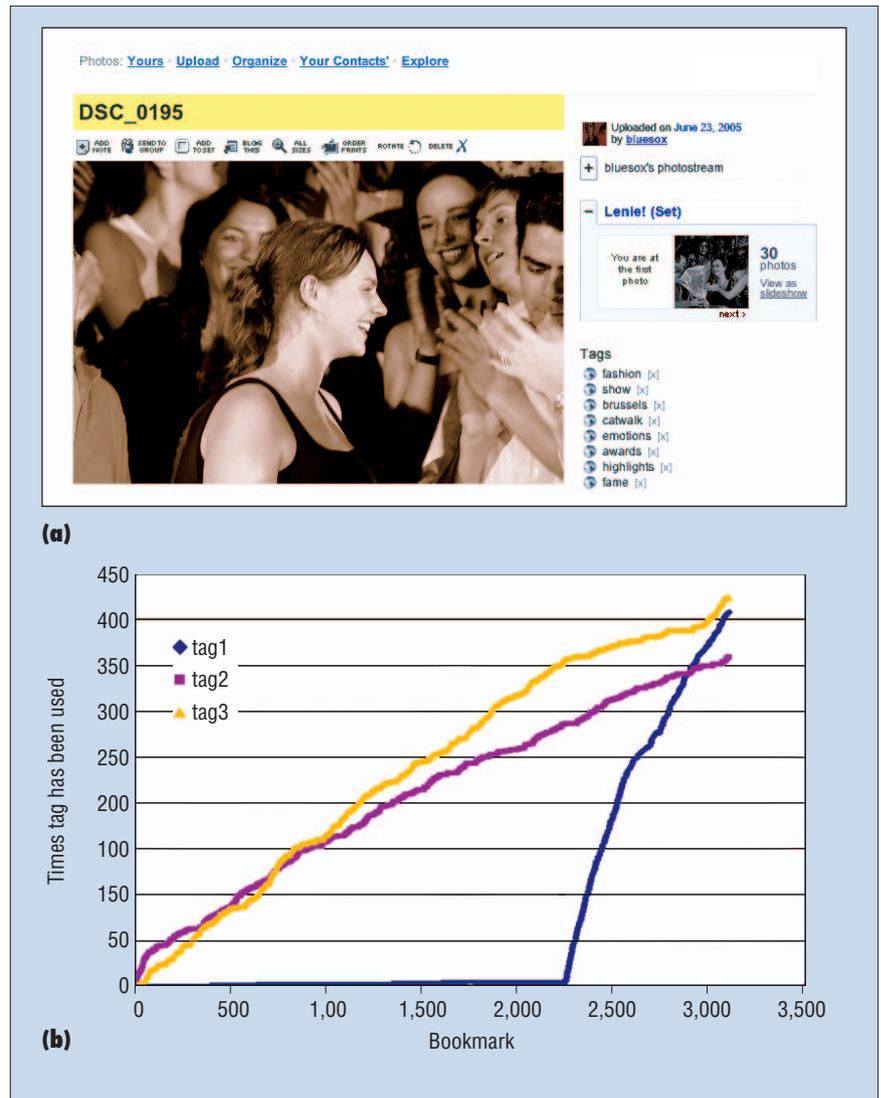


Figure 5. Collaborative tagging: (a) a screenshot of Flickr, a popular tagging site. Users tagged this picture with fashion, show, brussels, catwalk, emotions, awards, highlights, and fame; (b) a graph showing the frequency of tags referring to a Web site in del.icio.us. We see two tags stably established (even though this tag's user population grows exponentially) and then a third, new tag takes over, propagating like a virus in the population.

guage Origins," *Evolution of Communication*, vol. 1, no. 1, 1997, pp. 1–34.

10. A. Cangelosi and D. Parisi, "The Emergence of a 'Language' in an Evolving Population of Neural Networks," *Connection Science*, vol. 10, no. 2, 1998, pp. 83–97.
11. K. Smith, S. Kirby, and H. Brighton, "Iterated Learning: A Framework for the Emergence of Language," *Artificial Life*, vol. 9, no. 4, 2003, pp. 371–386.
12. L. Steels and F. Kaplan, "Bootstrapping Grounded Word Semantics," *Linguistic Evo-*

lution through Language Acquisition: Formal and Computational Models, T. Briscoe, ed., Cambridge Univ. Press, 2002.

13. L. Steels and T. Belpaeme, "Coordinating Perceptually Grounded Categories through Language: A Case Study for Colour," *Behavioral and Brain Sciences*, vol. 28, no. 4, 2005, pp. 469–529.
14. P.-Y. Oudeyer, "The Self-Organization of Speech Sounds," *J. Theoretical Biology*, vol. 233, no. 3, 2005, pp. 435–449.
15. L. Steels, "Constructivist Development of

The Author



Luc Steels is a professor of computer science at the Vrije Universiteit Brussel and director of the Sony Computer Science Laboratory in Paris. His research interests include artificial intelligence, developing the foundations of semiotic dynamics, and fluid-construction grammars. He is a member of the AAAI, IEEE, and ACL and a fellow of the European AI Society (ECCAI). Contact him at Vrije Universiteit Brussel, AI Lab, 10G-725, Pleinlaan 2, B-1050 Brussels, Belgium; steels@arti.vub.ac.be; <http://arti.vub.ac.be/~steels>.

Grounded Construction Grammars," *Proc. Ann. Meeting Assoc. for Computational Linguistics Conf.*, Assoc. for Computational Linguistics, 2004, pp. 9–19.

16. K. Wagner et al., "Progress in the Simulation of Emergent Communication and Language," *Adaptive Behavior*, vol. 11, no. 1, 2003, pp. 37–69.

17. J. Minett and W. Wang, *Language Acquisition*,

Change and Emergence: Essays in Evolutionary Linguistics, City Univ. of Hong Kong Press, 2005.

18. A. Baronchelli et al., "Sharp Transition towards Shared Vocabularies in Multi-Agent Systems," 2005; <http://arxiv.org/pdf/physics/0509075>.

19. D.J. Watts and S.H. Strogatz, "Collective Dynamics of 'Small-World' Networks,"

Nature, vol. 393, no. 6684, 1998, pp. 440–442.

20. A. Baronchelli et al., "Topology Induced Coarsening in Language Games," *Physical Rev. E*, vol. 73, no. 1, 2006, article 015105; <http://arxiv.org/abs/physics/0512045>.

21. L. Steels, "Language Games for Autonomous Robots," *IEEE Intelligent Systems*, vol. 16, no. 5, 2001.

22. L. Steels and M. Loetzsch, "Perspective Alignment in Spatial Language," *Spatial Language in Dialogue*, K. Coventry, J. Bateman, and T. Tenbrink, eds., Oxford Univ. Press, 2006.

23. S.A. Golder and B.A. Huberman, "The Structure of Collaborative Tagging Systems," *J. Information Science*, vol. 32, no. 2, 2006; <http://arxiv.org/pdf/cs.DL/0508082>.

For more information on this or any other computing topic, please visit our Digital Library at www.computer.org/publications/dlib.



JULY 13–15, 2006

DARTMOUTH COLLEGE
ARTIFICIAL INTELLIGENCE CONFERENCE

THE NEXT FIFTY YEARS

Commemorating the 1956 founding at
Dartmouth College of AI as a research discipline

With Original 1956
Conference Participants:

- John McCarthy (Stanford)
- Marvin Minsky (MIT)
- Ray Solomonoff (London)
- Oliver Selfridge (MIT)

Visit website for complete program

Explore the Future of
Artificial Intelligence with
Today's Foremost Intelligences

July 13-15, 2006

prior to the AAAI national annual meeting in Boston, MA

www.dartmouth.edu/~ai50



ai50@dartmouth.edu



603-646-8172