Operations on Multimodal Records: Towards a Computational Cognitive Linguistics

Andrzej Buller [1]

TR-95-027

June 1995

Abstract

The report discusses a cognitive model in which a key concept is Multimodal Record (MMR)--an organized aggregate of transcripts of signals representing all information an Agent continuously acquired for a certain period of time. The MMR consists of a video track, sound track, and a number of tracks containing transcripts of the values of temperature, pressure, etc., as well as transcripts of states of the Agent's internal structure. Three basic operations on MMRs, i.e. multimodal difference (m-), multimodal union (m+) and multimodal intersection (m*), to be performed using neural network has been introduced. Based on the operations one can explain and/or implement a number of psycho-linguistic phenomena. MMR may be considered as a computable form of Image Scheme--the basic concept of Lakoff-Langacker Cognitive Grammar. Hence, the proposed model seems to be a bridge over the gap between the non-computational Cognitive Linguistics and an applied neurocomputing. Moreover, it may be considered as a step towards a unified symbolic-connectionist paradigm.

Key words: Cognitive Grammar, Neurocomputing, Language acquisition

[1] The author is a Senior Fulbright Scholar on leave from the Faculty of Electronics, Technical University of Gdańsk, Gabriela Narutowicza 11/12, 80-952 Gdańsk, Poland, e-mail: buller@sunrise.pg.gda.pl

Abbreviations:

CL - Cognitive Linguistics
LTM - Long Term Memory
MMR - Multimodal Record
RAT - Running Agent
WM - Working Memory

1. Introduction

"Instead of trying to produce a programme to simulate the adult mind, why not rather try to produce one which simulates the child's? If this were then subjected to an appropriate course of education one would obtain the adult brain" wrote Alan Turing in 1950 anticipating that the challenging work would be completed in 50 years (Turing, 1950). Todays, 5 years before the dead-line proclaimed by Turing, one may observe that his visionary guidelines remain unquestionable. One can suppose, that the chance of meeting the dead-line has been strongly increased owing to the boost of popularity of Connectionism, Cognitive Linguistics (CL) and Society-of-Mind triggered in mid-80's by the publication of the works by Rumelhart et al. (1986), Lakoff (1987), and Minsky (1986). Indeed, the most ambitious projects oriented to a building of a thinking machine, for example, the L_0 project conducted at the ICSI, Berkeley (Feldman et al. 1993), the DETE project at the UCLA (Nenov and Dyer 1994), or the Cog project at MIT (Brooks and Stein 1993; Ferrel 1995), are strongly inspired by connectionism and CL. The Cog project employs also the basic concept of the Society-of-Mind.

For decades traditional linguistics has considered human linguistic behavior as a symbolic structure processing. The CL emerged as an opposition to the approach. While the traditional linguistics concentrates mostly on syntax, the CL deals with semantics, metaphors, and whole psychological, physiological, and social context. The key concept in CL is Image Scheme which became a basis of Cognitive Grammar (Langacker 1991). Although Lakoff (1995) promotes connectionist approach, the major obstacle in the way of an integration of connectionism and CL is fact that the later does not provide formal description of its basic concepts.

The approach to cognitive modelling presented in this report tries to overcome the obstacle. It integrates connectionist, CL-based, and Society-of-Mind-based approaches. Since human mind has to operate on uncertain data and is seldom required to produce precise solution, the proposed model adopts Soft Computing--a synthesis of Fuzzy Calculus, Neurocomputing, and Genetic Algorithms (Zadeh 1994)--as its basic tool-set.

2. Basic Concepts

Let us consider the scheme of an intelligent Agent consisting of Working Memory (WM), Long-Term Memory (LTM), Perceptive Block, and Executive Block (Fig.1). The scheme would look trivial, if it did not introduce a concept of mobile piece of information called Running Agent (RAT).

2.1. Running Agent (RAT)

RAT is is an 'agent inside Agent'. In other words the considered intelligent Agent has in its 'head' a society of disembodied, but mobile agents called 'Running Agents'. RAT is an aggregate of two separate pieces of data: (1) a strategy of its own behavior in the WM, and (2) an information to be processed in co-operation with other RATs inhabiting the WM. As RAT is assumed to be a mobile entity, one may consider it as a kind of computer virus. A typical RAT's task is: (i) to go to the Perceptive Block or to the LTM and get there a Multimodal Record (MMR), then (ii) to try to go to the LTM or the Executive Block respectively to deliver the MMR there. During the navigation in the WM a given RAT may encounter an other RAT and, depending on particular situation, either kill it or mate it, which may result in a birth of a new RAT carrying an MMR being a conclusion from the MMRs carried by its parents. Hence, a solution of a particular problem is a result of a number of independent RAT-RAT interactions which happen simultaneously in several places of the WM. An efficiency of the Society-of-RATs-based information processing increases owing to a certain evolutionary mechanism changing strategies of RATs' behavior.

2.2. Multimodal Record (MMR)

MMR is an organized aggregate of transcripts of signals representing all information the Agent continuously acquired for a certain period of time. The MMR structure consists of a video track, sound track, and a number of tracks containing transcripts of the values of temperature, pressure, etc., as well as transcripts of internal states of the Agent's internal structure. If a given MMR comes immediately from the Perceptive Block its video track contains more or less distorted images of real scenes. Processed MMRs may contain in their video tracks 'generalized' images or pointers to appropriate regions of the LTM where representatives of a family of images is stored. Hence, MMRs may be considered as a computable form of Image Schemata being used in the Lakoff-Langacker Cognitive Grammar (Buller 1992a)

2.3. Working Memory

An Efficient processing of data containing images seems to require connectionist way of computation. In the discussed model RATs are represented by patterns, while the WM is a recurrent neural network called Neural Screen (Buller 1992b; Buller 1993). However, unlike traditional connectionis, such pattern wanders unchanged in the network until it reaches one of the devices connected to WM or encounters other pattern representing a RAT. Hence, the Neural

Screen, as an implementation of WM, is to facilitate RAT's motion and RAT-RAT interactions. Unlike the most popular cognitive models, as for example the Baddeley-Hitch model (Baddeley 1995), the discussed WM does not deal with associating of visual perceptions and perceived sounds. The idea of MMR provides a convenient representation for association--simply, when two or more sensations are located in different tracks of a given MMR, this means that they are associated in time. Thus, the Perceptive Block provides associated data, while the WM facilitates their dissemination and processing.

3. Operations on MMRs

Information processing in the WM is a production of new MMRs which contain conclusions from records stored in LTM and from records produced by the Perceptive Block. Let us introduce the three basic set-theoretic operations for MMRs: multimodal difference (m-), multimodal union (m+), and multimodal intersection (m*). In order to discuss the role of the operations based on a sample cognitive process, let us assume that, as a result of a teaching the child-Agent, in a certain region of its LTM a number of MMRs <V0, S0>, where V0 - an image of a dog in the video track, S0 - a voice saing "dog" in the sound track is stored, while in an other separate region of LTM a number of MMRs <V1, S1>, where V1 - an image of a ball in the video track, S1 - a voice saying "ball" is stored. After a period of time the child sees a different dog. In such case the Perceptive Block produces a number of MMRs <V2, null>, where V2 - an image of the different dog, and injects the MMRs into the WM.

3.1. Multimodal difference (m-)

Every RAT carrying <V2, null> is able to knock some MMRs having something similar to V1 on their video tracks out of the LTM. Hence, in a while the WM is full of RATs carrying <V0, S0> and RATs carrying <V2, null>. An encounter of a RAT carrying <V0, S0> with a RAT carrying <V2, null>, will result in production of a new RAT carrying the MMR <null, S0> = <V0, S0> m- <V2, null>, with the probability dependent on the degree of similarity between V0 and V2 (see Fig.2). If the Executive Block is sensitive for MMRs having empty video track and non-empty sound track, the Agent, like a human child seeing a dog, will spontaneously pronounce the word "dog".

Let us assume that during a more advanced teaching, a teacher shows a dog, says "What's this?", and then immediately answers the question. Owing this, a number of MMRs <V0, S3S0>, where S3 - the voice saying "What's this?, will be stored in the LTM. When the teacher shows a different dog and asks "What's this?", the Perceptive Block will produce a number of MMRs <V2, S3> and inject them into the WM. A RAT carrying <V2, S3> can knock <V0, S0>s, as well as <V0, S3S0>s out of the LTM. Hence, owing to appropriate RAT-RAT interactions a number of RATs carrying <null, S0> = <V0, S3S0> m- <V2, S3> is born (see Fig.3.), which results in an activation of the Executive Block and a pronunciation of the word "dog" as a reaction to the visual perception of the animal accompanied by the question "What's this?" interpreted by the child-Agent as an indivisible sound effect.

3.2. Multimodal union (m+)

Once the teacher shows a different ball and asks "What's this?". In such case the Perceptive Block produces MMRs <V3, S3>, where V3 - an image of a different ball, S3 - the voice saying "What's this?". A RAT carrying <V3, S3> can knock some <V1, S1>s (based on the similarity between V1 and V3), as well as some <V2, S3S0>s (based on the common component S3) out of the LTM. However, in the population of MMRs knocked out, the number of <V1, S1>s should be greater than the number of <V2, S3S0>s. Hence, the RAT-RAT interaction will mostly result in the production of RATs carrying <V4, S3S1> = <V1, S1> m+ <V3, S3>, where V4 - a "generalized" image of a ball (see Fig.4) or a video-pointer to the LTM region in which MMRs containing V1s and V3s are stored. Owing to RAT motion, also <V4, S3S1>s will join the LTM region of V1s and V3s.

3.3. Multimodal intersection (m*)

An interaction between a RAT carrying <V3, S3> and a RAT carrying <V1, S1> may result in the production of <V5, null> = <V3, S3> m* <V1, S1> (see Fig.5). An so, the MMR <null, S1> able to make the Executive Block saying "ball" may be obtained either as <V4, S3S1> m- <V3, S3> or as <V1, S1> m- <V5, null>. Note that the Agent is able to answer the question "What's this?" in reference to ball in spite the fact that in has never been taught the sequence "What's this? ball" as it took place in case of dog. Moreover, if the teacher shows a ball for the second time and asks "What's this?", the Agent, having the <V4, S3S1> in its LTM will be able to provide an answer considerably sooner. The multimodal intersection may be also used for a removal not-essential elements from a perceived scene, which could be interpreted as a focusing of attention.

4. Machine which "wants" to know more

A continuous learning causes a gradual loading of the LTM. Owing to assumed properties of the LTM, an MMR carried by a RAT is located in a region in which MMRs similar to the incoming one are stored. Knocking MMRs out of the LTM depends on a degree of similarity between informational contents of stored MMRs and the knocking MMR. Based on the operations m-, m+, and m* the Agent learns answering more and more difficult questions about, for example, spatial relationships between perceived objects or spatio-temporal relationships between events. It also learns deriving conclusions from spoken stories. It is essential that the Agent also memorizes internal events which took place in its WM. However, it cannot talk about them.

Let us assume that an complete MMR has lower "energy" than incomplete one, and that the system WM-LTM is designed to always try to decrease its "total energy". Let also assume that the LTM has memorized a number of scenarios containing the question "What's this?" followed by a correct answer. In case of perception of an image Vk such that in the LTM there is nothing resembling the Vk, a new region of LTM is open to store the MMRs <Vk, null>. The presence of such incomplete MMRs in the LTM means increased "energy". Hence, the WM-LTM system tries to solve the problem of the "energy" reduction. It is believed that it will be able to generate a scenario in which the Agent says "What's this?".

5. RAT Farm

Evolution of the society of RATs takes place owing to a coerced two-phase cycle of mating. In the first mating phase the RATs mate to perform the cross-over of their strategies of behavior. Then, after optional mutation of the strategies the society of RATs is divided into groups following particular strategies. In the second mating phase each RAT searches of its mate to produce a conclusion from the MMRs it and its mate carries. In this phase there is no inter-group mating. After the end of this period the least efficient groups are eliminated and the entire cycle repeats. Owing to the mechanism, the RATs bred in the system become more and more intelligent, which, together with the development of the content and internal structure of LTM, results in an increasing knowledgeand intelligence of the Agent's artificial brain one could call RAT Farm. Now, let us discuss the way the RAT Farm inhabitants co-operate in solving a fuzzy problem.

Fuzziness is represented as a co-existence of contradictory statements associated with the same sensations. For example, let us assume that for a period of time the RAT Farm's Perceptive Block has produced p MMRs containing a given temperature T associated with the statement "too hot" and q MMRs containing the same temperature associated with the statement "not too hot", and that the records have been memorized in LTM. Although the value p/(p+q) may be interpreted as a membership of the temperature T to the fuzzy set "too hot", the discussed paradigm does not employ the classic, set-theoretic formulas in fuzzy inferencing. A sensation of, for example, the previously considered temperature T causes a knocking of p' MMRs containing the statement "too hot" and q' MMRs containing the statement "not too hot" out of the LTM, where p'/q' approximately equals p/q. RATs carrying the contradictory statements will compete in search of mates carrying MMRs containing the statements associated with pointers to appropriate actions. If, say, the RATs being adherents of the opinion that it is "too hot" win the competition, an alarm action will be raised. A crisp (non-fuzzy) statement is a special case of fuzzy statement for which q'/p' is very small. Some results of simulation of fuzzy calculus in such a way has been described in (Buller 1995ab).

6. Towards a unified symbolic-connectionist paradigm

The resurrection of connectionism in 80's evoked a great debate symbolists vs. connectionis. Indeed, neither pure connectionists nor pure symbolists managed to show a convincing way to a machine speaking intelligently in natural language. According to CL fundamentals, natural language is symbolic, but its acquisition is grounded on at least audio-visual perception. Surrounding world is full of symbols and people talk about the symbols. There are elegant models of cognitive processes based on logic (see, for example, McCarthy 1995). Also the molecular data processing using DNA strands (Adleman 1994), which may play an important role in an advanced Agent performance, seems to be just symbolic. Hence, although the proposed cognitive model is connectionis, it does not reject the notion of symbol, even on the level of internal processing. The RAT Farm project considers symbols as certain images stored in MMRs' video tracks. The images play a role of pointers to certain regions of LTM. And so, in an early phase of learning the string "2 + 3" may point to the region in which the text-book picture of two apples and three apples associated with the string. After more learning, the numerals and arithmetic operations become more abstract.

The model allows the situation that: (1) records wandering in neural network are processed by the network, while the network may be modifies according to the content of the records; (2) a record may knock other records out of the LTM when it has either similar content or a pointer to a particular region of the LTM; and (3) a record may contain images, sounds, smells, feeling of temperature, etc., as well as imperceptible data about internal structure of the WM-LTM system. Hence, the development of the RAT Farm may lead to such theory that a symbolic structure and a connectionist network are only two different facets of one unified representation (cf Buller 1994).

7. Concluding remarks

The cognitive model called RAT Farm shows that the idea ao the Society-of-Mind may help to build a bridge over the gap between connectionism and Cognitive Linguistics. The presented concept of Multimedia Record is a computable form of Image Schemata. Described examples of Multimedia Record processing explain selected psycho-linguistic phenomena, including a machine "will" to know more and a self-organization of its own learning. The processing integrates Fuzzy Calculus, Neurocomputing and Genetic Algorithms. It also combines symbolic and connectionist paradigm.

An implementation of an ultimate RAT Farm will be possible when an efficient neural network measuring a degree of similarity of short fragments of video pictures is built and successfully trained, which seems to not to be an extremely difficult goal. Such a network would be a major component of a processing module. A 2- or 3-dimensional mesh of such modules, where each module is connected with either its nearest neighbors, or with an Input/Output device, or with itself, would constitute a Working Memory for a brain-like computer working based on the discussed paradigm.

References

[1] Adleman L M (1994) "Molecular computation of solutions to combinatorial problems", Science, 11 Nov. 1994, vol.266 (5187), 1021-3.

[2] Baddeley A (1995) "Working Memory". In: Gazzaniga M S (ed.) The Cognitive Neurosciences, A Bradford Book/MIT Press: Cambridge, Mass., 755-764.

[3] Brooks R A, Stein L A (1993) "Building Brains for Bodies", A.I.Memo No. 1439, Massachusetts Institute of Technology, Artificial Intelligence Laboratory.

[4] Buller A (1992a) "MPM/LA: A Language Acquisition in a Massively Parallel Multimedia System", The Proceedings of First Singapore International Conference on Intelligent Systems (SPICIS'92), Singapore, 103-106.

[5] Buller A (1992b) Neural Screen: Towards a Connectionist Processing of Symbolic Knowledge, The Proceedings of First Singapore International Conference on Intelligent Systems (SPICIS'92), Singapore, 415-419.

[6] Buller A (1993) "Neural Screen and Intelligent Patterns", Proceedings of 1993 International Joint Conference on Neural Networks (IJCNN'93), Nagoya, 379-382.

[7] Buller A (1994) "Highly Distributed Processing and Sub-Neural Network", In: Carvallo M (ed.) Nature, Cognition, and System, Vol.3, 3-13, Kluwer Academic Publishers: Dodrecht/Boston/London.

[8] Buller A (1995a) "Society of Agents and Fuzzy Control", IASTED International Conference on Modelling and Simulation, Colombo, Sri Lanka, July 26-28, 1995.

[9] Buller A (1995b) "Fuzzy Inferencing: A Novel, Massively Parallel Approach", Technical Report, TR-95-025, International Computer Science Institute, Berkeley.

[10] Feldman J, Lakoff G, Bailey D, Narayanan S, Regier T, Stolcke A (1993) "L_0--The First Four Years" (manuscript to be published in AI Review, Vol.8, a special issue on Integration of Natural Language and Vision Processing).

[11] Ferrel C (1995) "Toward Building an Android: Recent Developments of the Cog Project", CS 298-10, AI/Vision/Robotics Seminar, May 9, University of California, Berkeley (abstract).

[12] Lakoff G (1987) Women, Fire, and Dangerous Things, University of Chicago Press: Chicago, Illinois.

[13] Lakoff G (1995) "Embodied Minds and Meanings". In: Baumgartner P, Payr S (eds.), Speaking Minds, Princeton University Press: Princeton, New Jersey, 115-129.

[14] Langacker R W (1991) Concept, Image, and Symbol: The Cognitive Basis of Grammar, Mouton de Gruyter: Berlin, New York.

[15] McCarthy J (1995) "Making Robots Conscious of their Mental States", Working Notes, AAAI Spring Symposium Series, Symposium: Representing Mental States and Mechanisms, Stanford University, March 27-29, 89-96.

[16] Minsky M (1986) The Society of Mind, Simon and Schuster: New York.

[17] Nenov V I, Dyer M G (1994) "Language Learning via Perceptual/Motor Association: A Massively Parallel Model", In: Kitano H, Hendler J A (eds.) Massively Parallel Artificial Intelligence, AAAI Press/The MIT Press: Menlo Park CA, Cambridge MA, London, 203-245.

[18] Rumelhart D E, McClelland J L and PDP Group (eds.) (1986) Parallel Distributed Processing. Exploration in the Microstructure of Cognition. A Bradford Book/MIT Press: Cambridge, Mass.; London.

[19] Turing A M (1950) "Computing machinery and intelligence", Mind, October, 59, s. 433-460.

[20] Zadeh L A (1994) "Soft Computing and Fuzzy Logic", IEEE Software, November, Vol. 11 (6), 48-56.

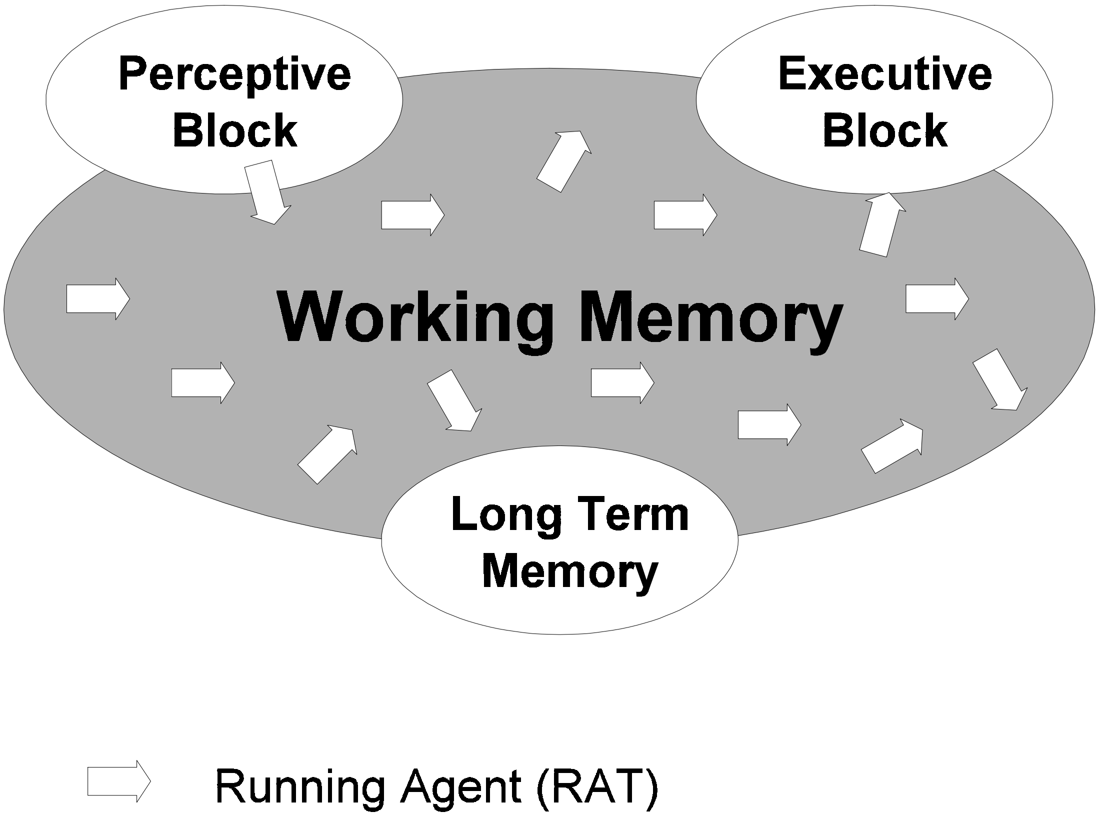

Fig.1. The proposed cognitive model of an intelligent Agent.
The mobile RATs are 'agents inside the Agent'.

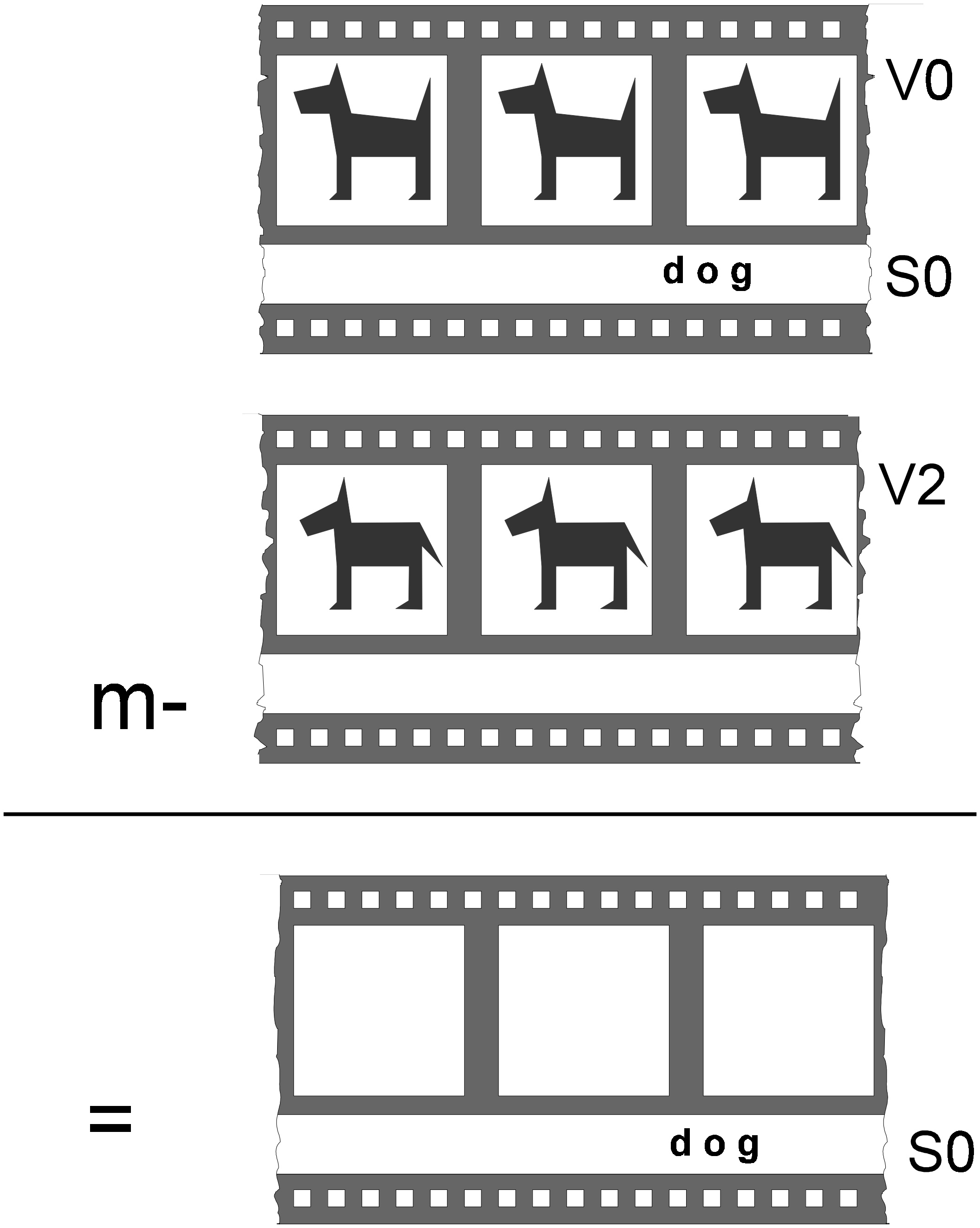

Fig.2. Multimodal difference.

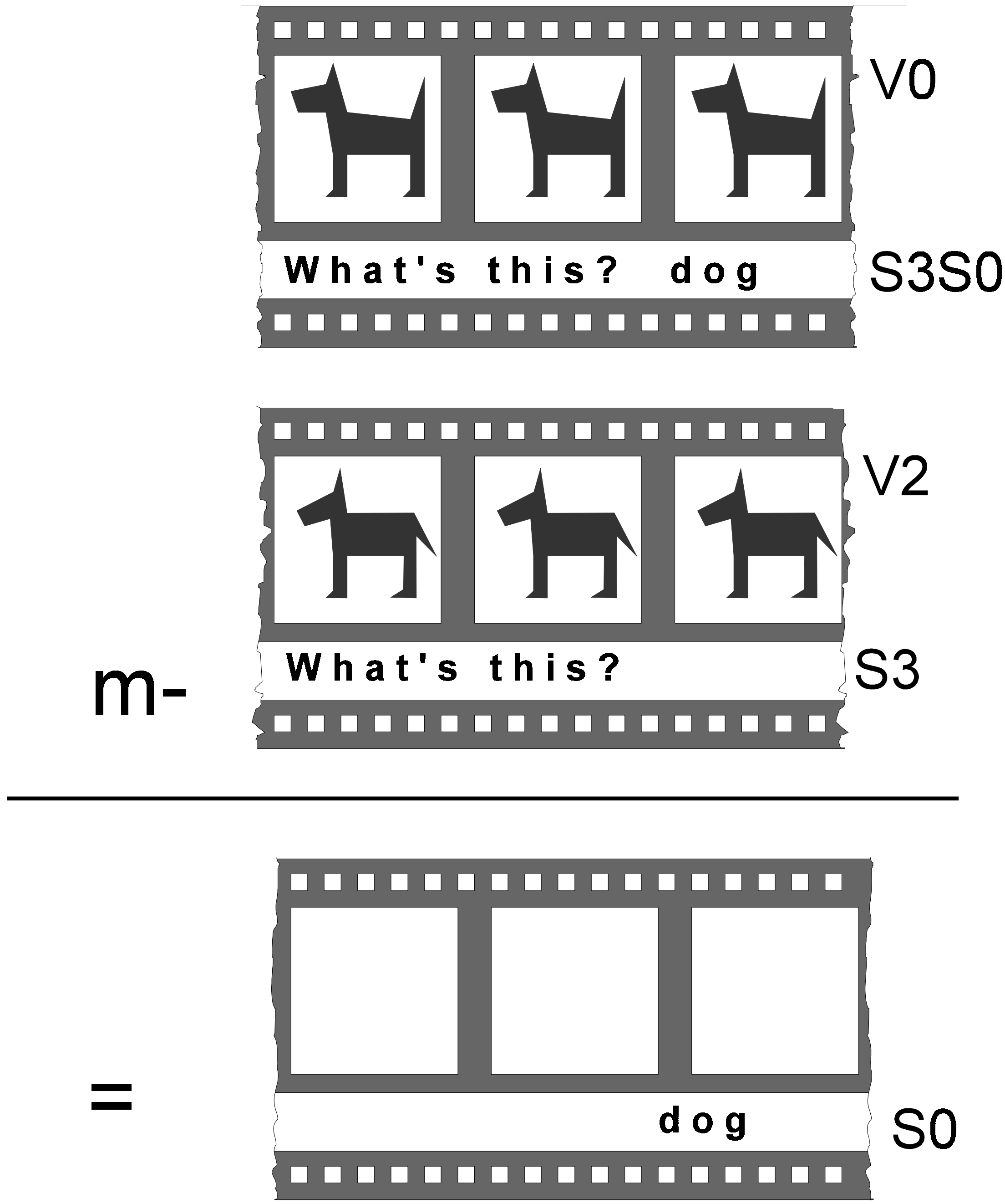

Fig.3. Multimodal difference. Another example.

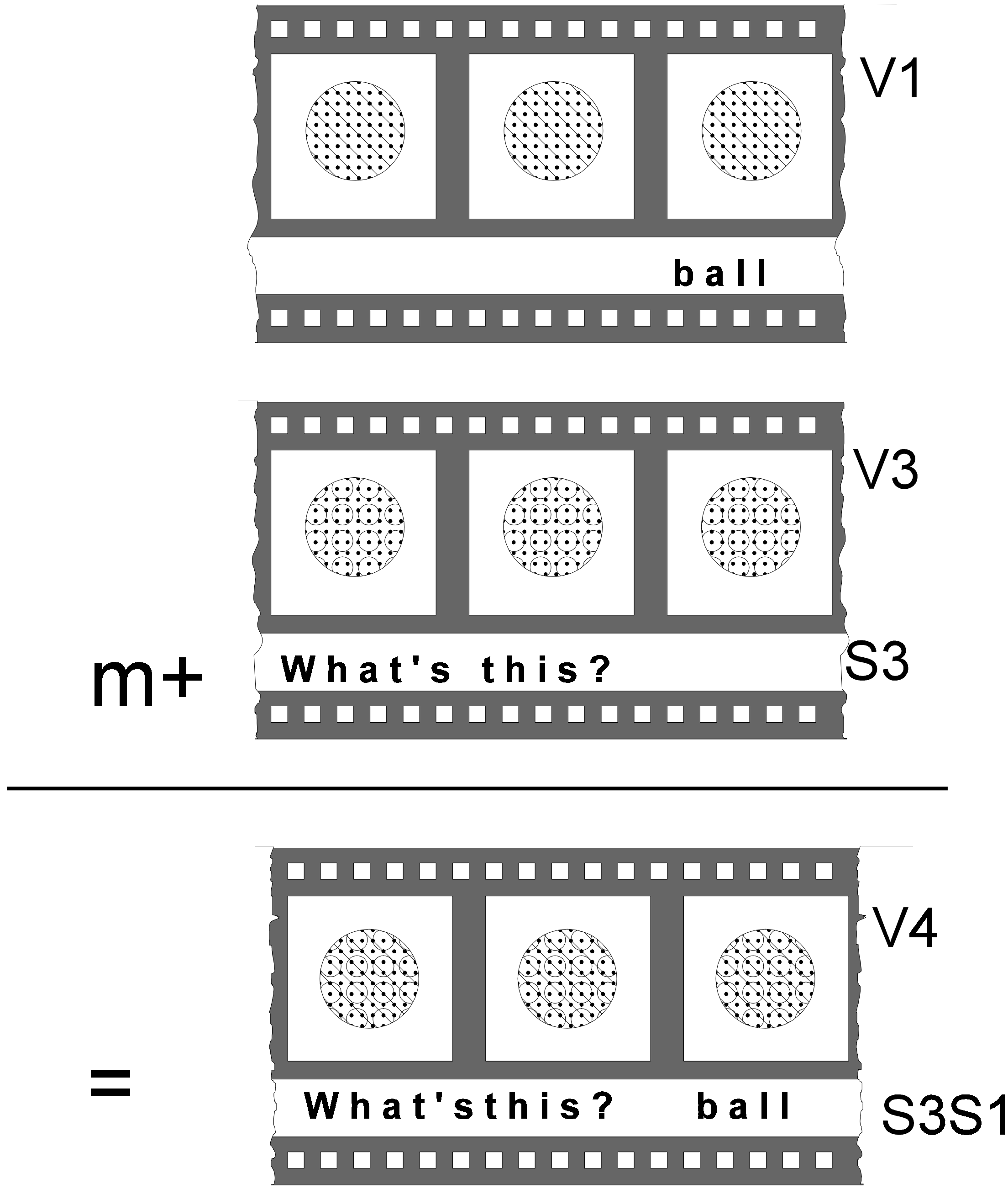

Fig.4. Multimodal union.

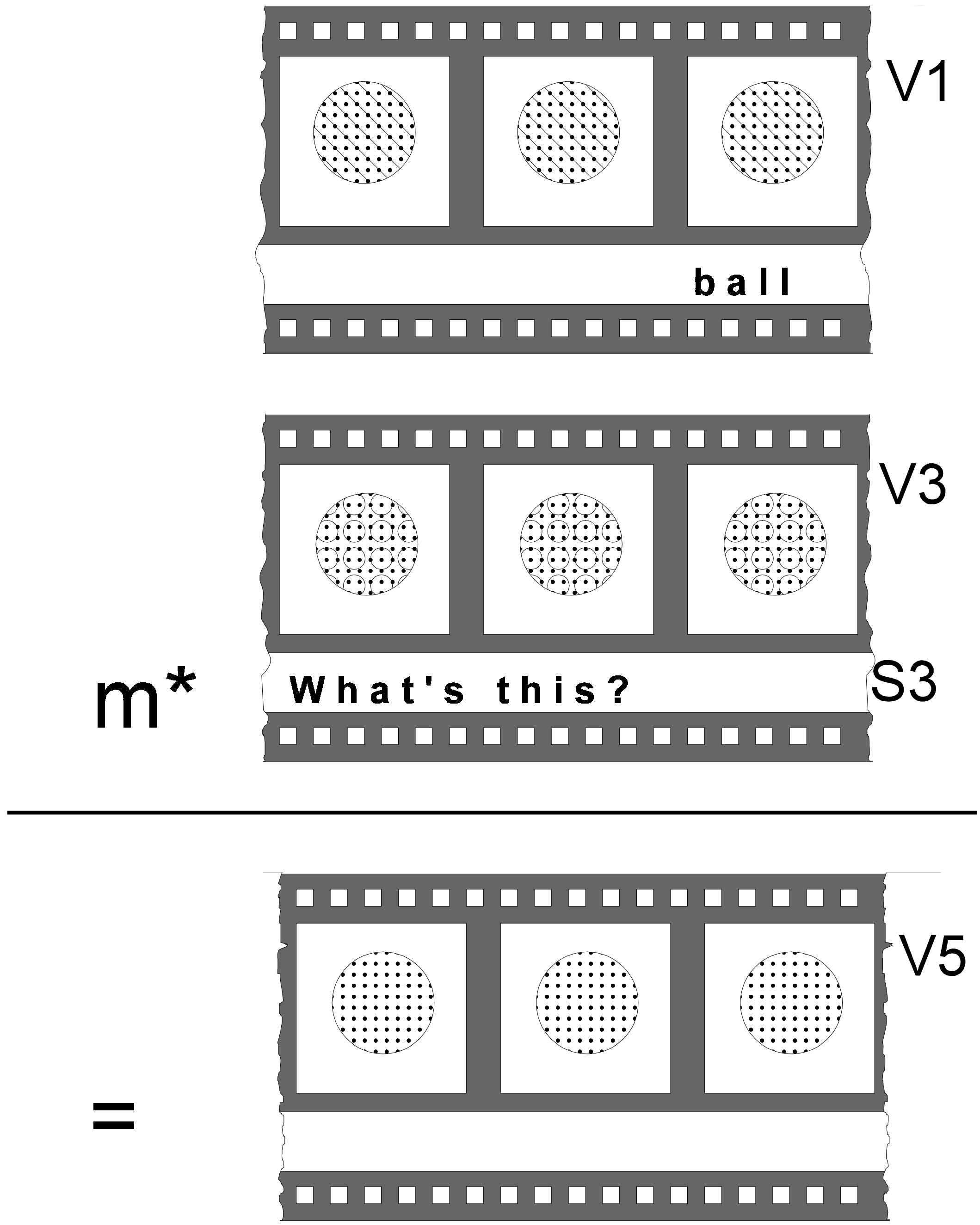

Fig.5. Multimodal intersection.