Typological and Computational Investigations of Spatial Perspective

Martin Loetzsch[1], Remi van Trijp[1], Luc Steels[1,2]

[1]Sony Computer Science Laboratory - Paris, 6, rue Amyot, 75005 Paris - France
[2]VUB AI Lab, Vrije Universiteit Brussels, Pleinlaan 2, 1050 Brussels - Belgium

Abstract. This paper is part of an ongoing research program to understand the cognitive and functional bases for the origins and evolution of spatial language. Following a cognitive-functional approach, we first investigate the cross-linguistic variety in spatial language, with special attention for spatial perspective. Based on this language-typological data, we hypothesize which cognitive mechanisms are needed to explain this variety and argue for an interdisciplinary approach to test these hypotheses. We then explain how experiments in artificial language evolution can contribute to that and give a concrete example.

1 Introduction

Traditionally, the 'language faculty' has been proposed to contain an innate and universal grammar, a view that has been defended by some influential thinkers such as Noam Chomsky and Steven Pinker [5,36]. A different stance is taken by cognitive (e.g. [25]) and functional linguistics (e.g. [16]), two related fields that study the general cognitive mechanisms that underly conceptualization and the functional pressures that explain differences in linguistic behaviour. Recently, a growing number of scientists have been working from a cognitive-pragmatic (or cognitive-functional) angle that combines the insights of both disciplines into a more complete language theory [35].

Our research subscribes to the cognitive-functional approach and therefore starts from the observation that language is a system that has both a functional dimension (linguistic behaviour) and a cognitive dimension (the biological nature or infrastructure of the language faculty), and that understanding language means understanding both dimensions and the correlations that exist between them. As argued in [34], the functional dimension of language can be directly observed in utterances and other forms of linguistic behaviour. The cognitive dimension, on the other hand, remains hidden in the 'black box' of the human mind. This means that the external use of language is the first and most important source of characterizing the inner workings of the mind, but at the same time the functional dimension of language can only be fully understood when more is known about the cognitive dimension, which implies that we have precise operational models of the information processing that goes on in cognition.

In this article, we first present a brief overview of spatial expressions across languages based on the literature on spatial language and our own study of ten languages[1]. Based on this data, we are able to formulate hypotheses on what cognitive mechanisms and operations are needed to make these kinds of expressions possible. We then present our research method, involving an experimental set-up with embodied communicative agents, that aims to go inside the black box and investigate these mechanisms. This research method is illustrated by an example experiment on spatial language, more specifically on the role of perspective reversal. Speakers of a language are able to use a different spatial perspective than their own in conceptualizing what to say and to explicitly mark this in language when needed. This is clearly present in road instructions (e.g. *Go straight ahead and leave the building to your left*), demonstratives, etc. Although there are obvious differences in how languages express perspective, there can be no doubt about the fact that they all have several ways to do so and that the speakers make abundant use of this facility. In the final section, we discuss the first results and directions for future research.

2 Some Observations of Spatial Language

Spatial language has already received a lot of attention in the past (see [45,46,25] for some groundbreaking research), but most of the studies on space grammars is based on familiar, western languages. This has led to some hasty conclusions, for instance that no language will have prepositions expressing specific shapes of objects such as *sprough*, meaning 'through a cigar-shaped object' [24], or that relative, anthropocentric spatial categories such as *left* versus *right* are universal or central to human spatial thinking [6]. However, recent cross-linguistic explorations have shown that human languages do not only vary in the syntactic structures that are employed to express spatial relations, but also in the set of semantic categories that are shared within a speech community (see [26,27] for a thorough investigation of 'relative' versus 'absolute' spatial expressions). In this section, we provide cross-linguistic examples of spatial expressions and look for cognitive-functional explanations that underly them.

We do not attempt to provide a complete semantic or syntactic typology of spatial expressions (some good reference works are [45,26,17]). Instead, we will give some relevant examples and try to answer the following three questions about spatial expressions that have been under serious debate in the past: (1) are people equipped with primitive spatial categories, (2) what is the semantic (and cognitive) nature of spatial categories across languages and (3) what grammatical items are used by languages to express spatial relations? We then present and define the dimension of spatial communication that we are especially interested in: spatial perspective.

[1] A complete overview of this study does not fit the purpose of this paper, so we made a selection of examples that illustrate our needs best. The ten languages and their reference grammars were: Alamblak [3], Dutch [15], Iraqw [33], Ket [50], Malayalam [13,1], Manam [28], North Marquesan [52], Páez [20], South-Eastern Pomo [32] and Zulu [4,10].

2.1 Spatial Categories

Many observers have assumed that spatial language may give us some insights into spatial cognition and the human mind in general. Starting from a nativist theory of language, or with the limited data of western languages, many scientists have claimed the existence of a universal set of spatial categories. This section suggests the opposite by showing a glimpse of the vast variety in how languages have decided to 'cut up' spatial relations.

Frames of Reference. One of the blackest flies in the ointment for everyone who defends the native view of language – whether they are talking about the innateness of syntax or the universality of semantic categories – is the fairly recent discovery of languages that do not use relative spatial categories such as *left* or *right* for locating objects in space. Instead, there are quite a few languages that prefer an 'absolute frame of reference' [26], comparable to spatial categories such as *North* and *South* in English.

The language Manam is a good example: its speakers live on a small island and their spatial language is dominated by two absolute directions (*ilau* 'towards the sea', and *auta* 'inland, towards the interior of the island'). All other directions are expressed by using these two. Across the seaward-inward line, Manam distinguishes between *ata* 'to one's left when facing the sea, and right when facing inland' and its opposite direction *awa*. When needed, the speakers of Manam can even combine these four directions for more precise indications. Adding the suffix *-lo* to the direction indicates motion. Thus the speakers of Manam wouldn't say something like 'The car is parked left of the tree' but rather something like 'The car is parked on the seaside of the tree'. Manam has a very complex and well developed range of spatial expressions, but all of them are based on this strong seaward-inland axis ([28], chapter 9).

(1) *áta* *i-sóaʔi*
left when facing seaward 3SG.RL-be located
'He is in the direction left when facing seaward.'

(2) *aúta-lo* *i-óro*
inland-MOTION 3SG.RL-go seaward
'He went in inland direction.'

The choice between different reference frames doesn't mean that they come with a universal set of spatial categories either. As documented in [26], the semantics of absolute spatial expressions can not be reduced to one single system of primitive categories. For example, some of these languages lack words that could be translated as 'left' or 'right' (as relative to a dominant axis). For example, North Marquesan only distinguishes an 'across' axis with respect to its dominant seaward-inland axis. To distinguish which side of this axis is referred to by the speaker, extra landmarks or place names have to be mentioned that help the hearer to find the right direction.

Also strongly related languages that make use of the relative frame of reference, such as English, Dutch and German, do not have spatial expressions that map easily from one to another. One can for instance easily observe the big differences between the formal and semantic properties of their spatial prepositions.

Specific versus Abstract. Traditionally, spatial categories have been regarded as carrying very abstract meaning. The parade example is the aforementioned prediction of Landau and Jackendoff that no language should have locatives like the hypothetical *sprough*, meaning 'through a cigar-shaped object' [24]. The classic counterexample is found in the Californian language Karuk that has a spatial suffix *-vara* with exactly this meaning ([26], p. 63).

When looking at the languages of the world, it becomes clear that the Karuk example is by no means an exception. For instance, South-Eastern Pomo has an inventory of 26 directional morphemes, most of which carry specific information such as the nature of the goal (water, land, etc.), the travel medium, deviations and changes in the motional state, etc. ([32], pp. 55–62, 79–91). The same language also contains motion verb roots that specify the presence or absence of the source of the motion, the relationship between the source and the referent that undergoes the motion (e.g. whether they touch each other or not), and the shape and orientation of the referent (long, standing, lying or vertical). The following example[2] shows how the semantics of the spatial markers specify the source object of the event:

(3) *líl*
into an enclosed space
-bò
crawling motion along a surface, into a long object horizontally
-t
IMPERF.
'He crawled into a tunnel.'

This example contains the directional morpheme *lil-* that specifies that the movement is into an enclosed space. This meaning is combined with the motion verb root *bo-* that specifies that the object is also long and horizontal, so the object that the speaker refers to can be translated as 'tunnel'.

Next to verb roots and suffixes, demonstratives too have been attested to give more specific information when they are used to indicate the relative distance from a referent to the deictic center. They can tell whether the *"referent is visible or out-of-sight, at a higher or lower elevation, uphill or downhill, upriver or downriver, or moving toward or away from the deictic center"* ([9], p. 170, also see [8]).

[2] The example came without glosses. Based on the information found in [32], we added them ourselves.

Example (4) shows an elevational marker in Alamblak, meaning that the referent is in a higher location than the speaker. Examples (5) and (6) show some demonstratives in Manam. As seen before, Manam bases its directions on the dominant seaward-inland axis and combines this with demonstratives so that the speaker can not only express the relative difference of a referent to herself, but also in which direction the referent should be situated. The speakers of Manam are able to express a four-way distance contrast with their demonstratives, ranging from nearby to far away and out of sight.

(4) *fëh-m-ko*
pig-3PL-up (higher than speaker)
'pigs up (there)'

(5) *áine* *éne* *i-tui=túi*
woman over there.across 3SG.RL-stand=REDUPL
'The woman is standing over there (left or right from the seaward-inland axis).'

(6) *i-alále* *enáwa-lo* *ʔába* *i-múle*
3SG.RL-go far over there-MOTION again 3SG.RL-return
enáta-lo
far over there-MOTION
'He went way over there (an out-of-sight place in the direction right when facing the sea) and then went back to way over there (an out-of-sight place in the opposite direction).'

In some cases, languages get even more specific. Alamblak has a number of locative words that can only be used with specific referents (e.g. trees, houses, canoes, large natural objects, etc.) ([3], p. 85). The most specific spatial expressions are place names, which are very often constructions that have become fixed names. Example (8) shows a noun phrase from Brabant (a Dutch dialect) that has become the name of a small forrest in the north of Belgium. We will not consider toponymy any further, but it should be noted that the use of place names is a very simple and effective way for people in a local community to talk about their environment.

(7) (a) *rawof*
'inside' (only with canoes)

(b) *mëfha*
'front' (only with canoes)

(8) *drei* *boom-ke-s-'* *berg-en*
three tree-DIMIN.-PL-GEN. mountain-PL.
Lit.: 'The mountains of the three small trees.'

Open-class vs Closed-class Subsystems. Traditionally, linguists have made the distinction between an 'open-class' or 'lexical' subsystem of language on the one hand, and a 'closed-class' or 'grammatical' subsystem on the other hand. As argued in [47], the closed-class subsystem determines conceptual structure and should therefore be the focus of research if one wants to investigate the spatial structuring in language. While this distinction is very useful for scientific purposes, it does not capture the complete picture of language.

The distinction between an open-class and closed-class subsystem has been conceived from a static view on language. However, languages are constantly changing and research in grammaticalization shows that closed subsystems are not as closed as they appear to be [49]. This has led to Paul Hopper's notion of 'emergent grammar' [18], that is, grammar is always on the move. The recent development of the collostructional analysis in corpus linguistics allows researchers to detect latent grammaticalization processes that can only be uncovered by looking at large amounts of data [44,30]. The following example (taken from [2], p. 163) shows how the speakers of Thai use the 'open-class' word *maa* 'come' in a serial verb construction to mark the destination of an event.

(9)	*thân*	*cà*	*bin*	*maa*	*krungthêep*
	he	will	fly	come	Bangkok

'He will fly to Bangkok.'

When it occurs in serial verb constructions, *maa* cannot be inflected independently for tense, mood, or aspect. A subsequent step in the grammaticalization process could be the re-interpretation of the verb as an adposition. It is widely attested that lexical verbs are a big source of grammaticalization for prepositions, case-markers and other grammatical items.

Thus when looking at actual language data, there seems to be no sharp distinction between lexical and grammatical items. This has been recognized and explicitly addressed by many studies in cognitive linguistics and construction grammar [22,14]. These theories represent linguistic knowledge as a continuum from the lexicon to syntax, which is an important observation for building a model that is in line with what is known about cognition.

Moreover, spatial relations can be expressed by virtually every grammatical item in language: motion verbs, cases (e.g. Finnish distinguishes between the interior locative cases inessive, elative and illative, and the exterior locative cases adessive, ablative and allative), spatial prepositions, adnominals, adverbial phrases, three-place locative constructions (e.g. Alamblak), demonstratives, etc.

2.2 Spatial Perspective

Our main research focus lies in **'spatial perspective'**, that is, how speakers express a scene as perceived by the visual system and how they are able to cope with the different angles from which the different speech participants observe the world. Our notion of spatial perspective more or less corresponds to what Talmy calls the 'perspective point' – the *"point within a scene at which one* *conceptually places one's 'mental eyes' to look out over the rest of the scene"* ([45], p. 217, see also examples (11–12) further down). The following examples give a clear illustration of spatial perspective.

(10) The car is parked on this side of the house.
(11) There are some houses in the valley.
(12) There is a house every now and then through the valley.
(13) The ball rolled from *my left* to *your right*.

In the first example, the speaker means that the car is located on the side of the house at which she is standing. In other words, the scene should be viewed from her spatial perspective. Examples (11) and (12) show how complex spatial perspective can be when it is combined with other cognitive mechanisms. In (11), the spatial perspective is stationary and the scene is viewed from a certain distance, whereas example (12) shows that the same valley can be conceptualized from a viewpoint in which one can see every individual house from up close, following a motion *through* the valley. The last example shows how there can be a **shift in spatial perspective**. Sentence (13) illustrates how the speaker explicitly marks from which viewpoints the hearer has to interpret the locative expressions in the utterance. The source of the ball movement is 'left of the speaker' as seen from the speaker's point-of-view. The target location of the roll-event contains information that should be interpreted from the hearer's own spatial perspective.

Spatial perspective is also shown in deictic markers such as pronouns or demonstratives. Example (14) gives three Japanese demonstratives, of which *sono* explicitly means that the referent should be located near the hearer ([23], cited from [9]). Finally, even if the speaker expresses a spatial relation from her own point-of-view, the hearer often has to perform egocentric perspective reversal to be able to interpret the utterance.

(14) *kono+INFL* *sono+INFL* *ano+INFL*
near speaker 'near hearer' 'away from speaker and hearer'

Perspective reversal also occurs when using landmarks to situate a referent or to indicate a direction, which is a complicated matter that also varies from language to language. Given the scope and space limits of this paper, we will not go into this topic now, but we will come back to it in the last section of this paper when we discuss the further steps in our research program. We kindly refer the interested reader to ([26], chapter 3) for a brief overview of the linguistic diversity regarding this subject.

2.3 Conclusions

Based on the cross-linguistic data of our own study and the results of other typological research, we can draw the following conclusions with respect to cognitive mechanisms needed for spatial language:

1. Language users must be able to impose a reference frame on their environment. A reference frame contains a point of view (perspective) from which the world is perceived, and local (i.e. temporary and viewpoint-dependent) or global landmarks. By default the perspective on the scene is the position of the speaker in the world, because the vision system directly produces perceptual features from this position.
2. There is strong evidence that there are no universal spatial categories: every language has its own way of cutting up the perceptual space. This implies that the language faculty should include cognitive mechanisms that allow a group of speakers to create new spatial categories. There are of course trends in languages because spatial categorisation is obviously constrained by the properties of the real world and our embodiment in that world. Spatial categories divide the perceptual continuum into discrete regions, such as left/right, front/back. Some categories are relational in the sense that they discretise the spatial relation between different objects located in the reference frame (as in "the ball left of the box").
3. There is strong evidence for a continuum from specific to abstract categories and from lexical to grammatical items (e.g. [7,11]). Moreover the examples given earlier make it abundantly clear that different languages make different choices with respect to what spatial categories or relations are lexicalised or grammaticalised. This implies that the language faculty must give language users the ability to lexicalise or grammaticalise spatial concepts, as opposed to support the usage of hard-coded lexical or grammatical constructions. If every language user has the capacity to invent their own categories and decide himself which ones to lexicalise or grammaticalise, then there is a risk of incoherence, so language users must also be able to negotiate with each other which linguistic conventions are to be commonly accepted by the group.
4. Finally, it is clear that language users are able to adopt another reference frame than their own. This implies that they are capable of egocentric perspective transformation (EPT), i.e. to compute what the world looks like from another perspective, particularly that of the other participant in the dialogue.

3 The Perspective Reversal Experiment

Psychologists and neuroscientists have made quite a lot of progress to identify cognitive mechanisms that are involved in the language faculty. For example, the capacity to perform egocentric perspective transformation has been shown to be universally present in normal humans [38] and possibly animals [31]. Neurological evidence has shown that it is carried out in the parietal-temporal-occipital junction which is active whenever its function is needed [51]. Egocentric perspective transformation is used in a variety of non-linguistic tasks, such as the prediction of the behaviour of others in navigation [19].

These studies typically identify that humans are capable of a certain cognitive task and where in the brain the processing necessary for this task might be performed, however they do not give a precise detailed operational model of exactly what kind of processing is needed, neither of the information structures that are required, how the information might be obtained by the cognitive agent, nor of the information transformations or the order in which they are executed. [This is like observing that humans are able to fight off bacteria and that the liver is involved in this process but without detailing exactly what metabolic pathways or biochemical processes are actually doing the work.] Today it is however possible to make such precise operational models and advances in Artificial Intelligence and robotics enable us to build sufficiently complex artificial 'agents' that contain implementations of these models and to test out whether they are adequate. This is precisely the research task that we are pursuing in our laboratory.

Our research methodology involves the following steps (see [40]):

1. Pick a feature of language.
2. Look at the linguistic coding of this feature in different languages.
3. Hypothesize which cognitive mechanisms and external factors (functional pressures) are necessary for the emergence of this particular feature.
4. Operationalize the mechanisms in computational processes and endow agents with these mechanisms.
5. Build a scenario of agent interaction, preferably embedded in some simulation of the world. This scenario and the virtual world have to pose the specific communicative challenges that trigger the need for the investigated language feature.
6. Perform a systematic series of simulations, demonstrating that the feature indeed emerges and that the cognitive mechanisms are in fact necessary. Ideally, this is shown by comparing simulations in which the agents do not have these mechanisms at their disposal to simulations in which they are endowed with them.

The remainder of this section gives a concrete example in which some of the cognitive mechanisms needed for spatial language have been worked out. The experiment is described in more detail in [43]. It features robots that roam around freely in an unconstrained office environment and play language games [42] about ball movement events (see figure 1). We consider this experiment only to be the first step, as we restrict spatial cognition for the time being to spatial categories only (not yet relations), and use a purely lexical language, even if there can be multiple words.

3.1 Embodiment

The robots are fully autonomous Sony Aibo ERS7 [12][3]. Based on software developed for robotic soccer [37], a real-time image processing system ([21], see left column of figure 2), probabilistic modeling techniques for the maintenance

[3] Main sensor: 208×160 pixel digital camera, 20 degrees of freedom, 400 MHz Mips processor, distance sensors, microphone, speakers, wireless communication.

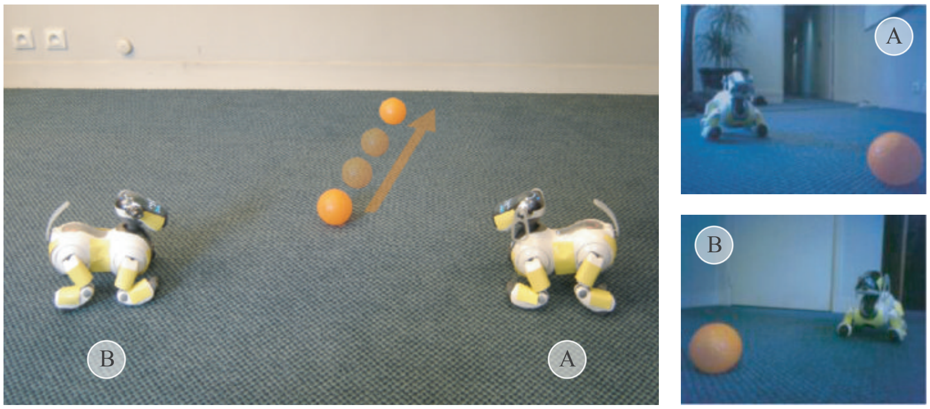


Fig. 1. Left: An example scene. Two Aibo robots (robot *A* and *B*) observe a ball movement and then describe the scene to each other. Right: The scene at the left seen through the built-in cameras of the two robots.

of a persistent, analog, and egocentric model of the ball and the other robot (see middle column of figure 2), object tracking, locomotion, and obstacle avoidance were built into the robots.

Behavior control programs [29] were made for coordination between robots. Both robots randomly walk around while avoiding obstacles. Each robot that sees both the ball and the other robot sends an acoustic signal. Robots continue with random exploration until a configuration is reached so that they can establish a joint attentional frame, in the sense of [48] (see below in section 3.2). When both robots are ready to observe the scene together, a human experimenter manually moves the ball. The begin and end point of the trajectory (see right column of figure 2) are recorded and sent to the language system via the wireless network. Continuous values on 12 channels are extracted from such descriptions and put into the world model of the agent. In order to be able to repeat the experiment with the same data in different experimental conditions (and in order to accelerate the process), we recorded about 250 such world models for both robots and used them later on in simulations.

This basic sensory processing achieves the first cognitive mechanism identified earlier, namely the ability to create a reference frame from the viewpoint of the agent with objects located within this frame. In this experiment, there are no global landmarks (although they could potentially be implemented), only local landmarks directly in the field of view of the agents.

We have also made a very concrete operational model of the egocentric perspective transformation that is clearly recognised as fundamental in spatial language (see figure 3). The model is implemented by taking the features (such as

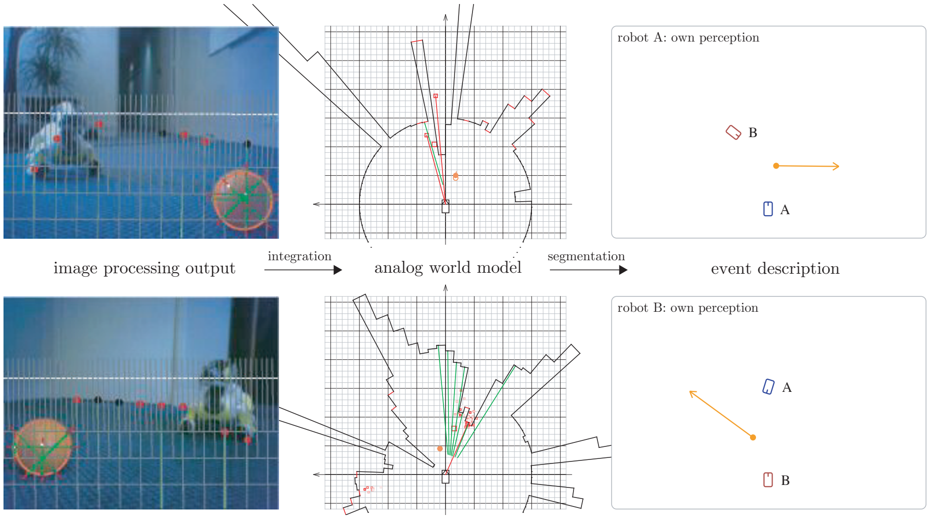


Fig. 2. From images to event descriptions. Left: Real-time model based image processing algorithms scan the camera image along a horizon-aligned grid to detect balls, other players and obstacles. An orange circle denotes a detected ball and black and red dots denote detected obstacles. Middle: The percepts from each camera image are integrated into an analog world model. Green lines denote obstacle percepts, red lines perceived positions of robots. Red squares are hypotheses for the position of the other robot. The filled red square is the estimated position of the other robot, the filled orange circle is the filtered position of the ball. The dark lines around the robot represent the filtered distances to obstacles. Right: Event descriptions extracted from the analog world model.

angle of movement, position of an object, etc.) and transforming them given the position of another object (such as the position of the other robot in the scene).

3.2 Language Games

A language game is a constrained routinised interaction between two agents. It involves two aspects. First a joint attention frame [48] needs to be established, which means that there must be a shared motivation, a shared communicative goal, and shared attention to the same object in the environment. This joint attention frame is part of the scripts that the robots follow in their interaction. Given a joint attention frame a verbal interaction can take place - which in this case is a description game. One agent describes to another one the most recent event that involves the orange ball. The description must not only be true but also distinctive. Agents then give each other feedback whether the interaction was successful or not.

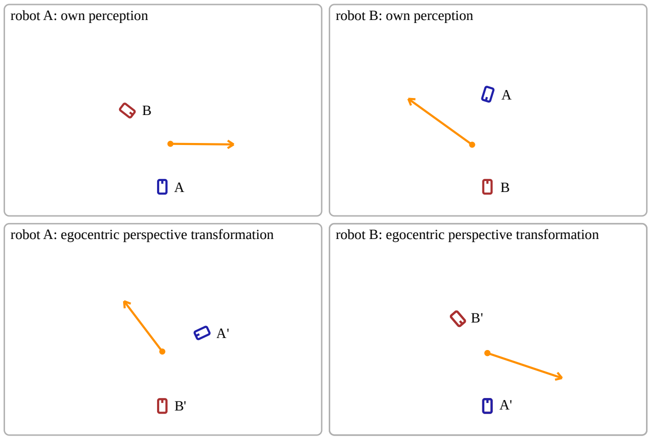
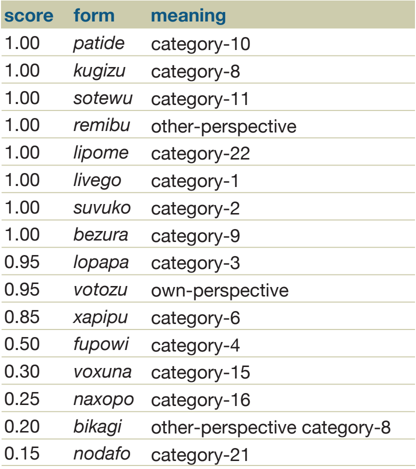

robot A: own perception
B
A
robot B: own perception
A
B
robot A: egocentric perspective transformation
A'
B'
robot B: egocentric perspective transformation
B'
A'

Fig. 3. Egocentric Perspective Transformation. Top row: The event from figure 1 as perceived by robots A and B. Bottom row: The result of egocentric perspective transformation. Both robots are able to construct a description of the scene as it would look like from the perceived position of the other robot.

Apart from the mechanisms to follow the script required to play a game, the agents are endowed with the other cognitive mechanisms that we identified earlier. The first one is the ability to use and create new spatial distinctions to discriminate the 'topic' (the current event) from a 'context' (the previous event). It is based on discrimination trees [39]. Perceptual channels are hierarchically divided into equally sized regions. For example the category `category-4` covers the interval `[0,0.5]` on channel `ball-y2`, meaning that the ball ends right. Whenever distinctive categories cannot be found, the agent extends his ontology by cutting up a perceptual channel into different regions, and progressively they develop enough categories to make all the distinctions that are needed in this domain. The present experiment does not (yet) endow agents with the ability to deal with relational categories.

Agents use not only their own perspective but also that of the hearer. If the discriminating category works from both perspectives, perspective does not need to be included in the meaning that the speaker is going to express (the perception of that feature of the scene is shared). Otherwise, the meaning must include information from which perspective the categorisation took place. Usually there is more than one way to conceptualize the scene. Categories are ranked based on saliency and score obtained from earlier success in the game. The description with the highest score is used further.

Agents need yet another mechanism, namely the ability to maintain a bi-directional inventory of meaning-form pairs and the ability to extend the inventory either because they need to express a new spatial relation or because they



score	form	meaning
1.00	*patide*	category-10
1.00	*kugizu*	category-8
1.00	*sotewu*	category-11
1.00	*remibu*	other-perspective
1.00	*lipome*	category-22
1.00	*livego*	category-1
1.00	*suvuko*	category-2
1.00	*bezura*	category-9
0.95	*lopapa*	category-3
0.95	*votozu*	own-perspective
0.85	*xapipu*	category-6
0.50	*fupowi*	category-4
0.30	*voxuna*	category-15
0.25	*naxopo*	category-16
0.20	*bikagi*	other-perspective category-8
0.15	*nodafo*	category-21

Fig. 4. The lexicon of agent 3 after 4412 games.

hear a new construction used by another agent. Agents invent new words by combining random syllables if needed.

A game is a success if the hearer knows all the words in the utterance and if the extracted meanings are true and discriminating for the current event. Everything else is a failure. Communicative success is the only measure that drives the coherence of perceptual categories and lexical items among the agents of a population. Each category and meaning-form association has a score that reflects its overall success in communication. After a successful game, the score of the lexical entries that were used for production or parsing is increased by 0.05. At the same time, the scores of competing lexical entries with the same form but different meanings are decreased by 0.05 (lateral inhibition). In case of a failure, the score of the involved items is decreased by 0.05.

3.3 Testing Different Configurations of Cognitive Mechanisms

The main advantage of computational modeling is that we can be very precise in terms of what information processing has to go on to achieve a particular function. But we can do even more because we can test different configurations of cognitive mechanisms to prove why they might have been adopted universally for human languages. This section illustrates this methodology showing that egocentric perspective transformation is not just a luxury which accidentally became used, but is highly useful for increasing the communicative success and decrease the cognitive efforts required by language users.

We tested the dynamics of the evolving communication system for four different configurations of the cognitive mechanisms described earlier. To be able to

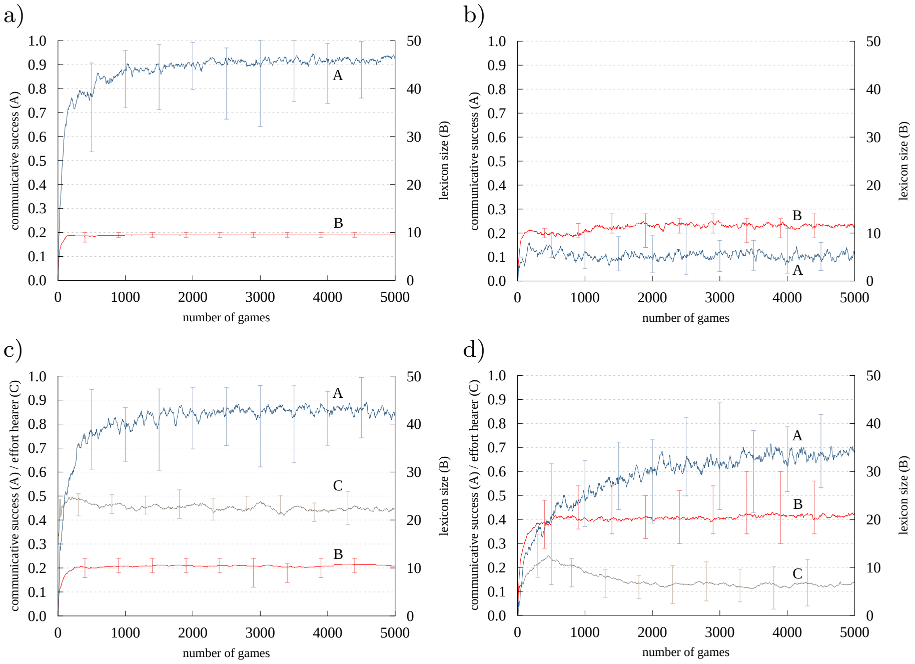


Fig. 5. Experimental results averaged over 10 runs of 5000 language games each in a population of 5 agents: Curve A shows communicative success (fraction of successful games in the last 100 interactions). B is the lexicon size averaged over all agents of the population. C is cognitive effort of the hearer (how often the hearer has to do an additional perspective transformation during interpretation). Experimental conditions: in a), both agents share the same perception and therefore don't have to do perspective reversal. In b), the agents view the scenes from different angles but don't do Egocentric Perspective Reversal. In c), the agents are capable of doing EPT but don't mark it in language. In d), perspective is also marked in language

compare the results, two events from the same set of 250 recorded world models were randomly selected for each interaction.

To show that the implemented cognitive mechanisms do indeed work, we first ran an experiment where both speaker and the hearer don't have the ability to do Egocentric Perspective Reversal but perceive the world from the same point of view (by artificially letting the hearer perceive the same scene descriptions as the speaker). As shown in figure 5a the agents reach more than 90% communicative success after about 1000 interactions. The average lexicon size stabilizes at around 10 words. If, as shown in figure 5b, the speaker and the hearer perceive scenes from different angles and if they are not able to do EPT, they are not able to establish a communication system. Communicative success does not exceed 15% and the average length of the non-aligned lexicons is about 12 words.

In a third configuration, the agents are able to use EPT for conceptualization. That means that the speaker conceptualizes the scene both from his own perspective and the perspective of the hearer. The more salient semantic description is then lexicalized. The hearer immediately adopts the speaker's perspective by performing an EPT on the own world model. If he cannot interpret the utterance in that world model, the own perspective is additionally tried. By doing that, the agents are able to self-organize a communication about ball movements although viewing the scene from different angles, as shown in figure 5c[4]. However, the hearer has to perform additional perspective transformations in interpretations. This is captured in the 'cognitive effort' curve (C) in figure 5c).

In the fourth configuration, the perspective chosen by the speaker is also marked in language. That means that a perspective indicator (`own-perspective` vs. `other-perspective`) is added to the list of predicates resulting from conceptualization. Note that there is no bias towards a specific way to lexicalize perspective marking. Instead, these perspective indicators are treated in the same way as any other predicate coming out of conceptualization. The fact that single perspective markers emerge is a side effect of the general lexicon process. Given this configuration, the cognitive effort for the hearer significantly drops (figure 5d) as the hearer immediately knows which perspective to use[5].

As the results show, egocentric perspective reversal is an essential prerequisite for communication situated in space. Remarkably, it takes only a few thousand interactions until the 5 agents align their conceptual and linguistic inventories. Whereas in the beginning word meanings tend to be more holistic, as for example the word *bikagi* in figure 4 holistically covers the meaning `other-perspective category-8`. Later on, agents start to generalize and perspective is lexicalized separately, e.g. *votozu* for `own-perspective` (figure 4). An example utterance looks like this:

(15) *fupowi* *remibu*
other-perspective category-4
'ends to your right'

Even though the evolved languages feature multi-word utterances and separate lexical perspective markers (example 15), they are not grammatical. Given this experimental setup, a purely lexical language covers all the communicative needs, but we have already started to experiment with a richer world model that contains opportunities for the agents to recruit additional cognitive mechanisms, including multiple landmarks so that spatial relations become an additional resource in the language game.

[4] The fact that the communicative success is slightly less than in the first condition (figure 5a) is the result of noisy perception.

[5] Note that communicative success does not converge as fast as in conditions a) and c) because the learning problem is more difficult.

4 Conclusions and Further Research

This paper illustrates how different subfields of cognitive science can interact to build a comprehensive theory of spatial language. From linguistics, we get observations of the kinds of how spatial categories and relations get expressed and of the flexibility that is apparently required. Given the evidence, it is clear that spatial concepts are not genetically hard-coded and neither is there a simple universal mapping from spatial cognition to spatial language. Instead there is a lot of variety which necessitates that language users must be seen as creatively expanding and negotiating their repertoire of spatial concepts and their linguistic conventions.

Although psychology and neuroscience can give us hints on the kinds of cognitive capabilities humans have and where approximately in the brain the information processing to achieve these capabilities might be located, it is only by making concrete detailed operational models that we can actually understand how spatial cognition and language is possible: in contrast to earlier computational work, where spatial cognition, lexicons and grammar are implemented by hand. This paper reported breakthrough experiments showing that the current state of the art in AI and robotics is sufficiently advanced to carry out highly non-trivial experiments that test operational models of spatial language evolution, in other words where agents invent and negotiate a spatial language of their own making. In addition, we demonstrated that egocentric perspective transformation is beneficial for establishing a communication system among agents that are able to see scenes from different spatial perspectives.

Acknowledgements. Our experiments rest on the Fluid Construction Grammar framework [41], to which Nicolas Neubauer and Joachim De Beule have made major contributions. The authors also thank the members of the "GermanTeam" for providing their robot soccer software. This research was funded and carried out at the Sony Computer Science Laboratory in Paris with additional funding from the EU FET ECAgents Project IST-1940.

References

1. M. S. Andronov. *A Grammar of the Malayalam Language in Historical Treatment.* Beitrage zur Kenntnis südasiatischer Sprachen und Literaturen. Harrasowitz Verlag, Wiesbaden, 1996.
2. B. J. Blake. *Case.* Cambridge Textbook in Linguistics. Cambridge University Press, Cambridge, 1994.
3. L. P. Bruce. *The Alamblak Language of Papua New guinea (East Sepik).* Pacific Linguistics Series C 81. Australian National University, Canberra, 1984.
4. N. N. Canonici. *Zulu Grammatical Structure.* university of Natal - Durban, Natal, 1995.
5. N. Chomsky. *Aspects of the Theory of Syntax.* MIT Press, Cambridge, MA, 1965.
6. H. Clark. Space, time, semantics and the child. In T. Moore, editor, *Cognitive Development and the Acquisition of Language*, pages 28–64. Academic Press, New York, 1973.

7. W. Croft. *Radical Construction Grammar: Syntactic Theory in Typological Perspective.* Oxford University Press, Oxford, 2001.
8. H. Diessel. *Demonstratives: Form, Function, and Grammaticalization.* Typological Studies in Language 42. John Benjamins, Amsterdam, 1999.
9. H. Diessel. Distance contrasts in demonstratives. In M. Haspelmath, M. S. Dryer, D. Gil, and B. Comrie, editors, *The World Atlas of Language Structures*, pages 170–173. Oxford University Press, Oxford, 2005.
10. C. M. Doke. *Textbook of Zulu Grammar.* Maskew Miller Longman, Cape Town, 1992. Previously published in 1927 by the University of Witwatersrand.
11. M. S. Dryer. Are grammatical relations universal? In J. Bybee, J. Haiman, and S. Thompson, editors, *Essays on Language Function and Language Type: Dedicated to T. Givon*, pages 115–143. John Benjamins, Amsterdam, 1997.
12. M. Fujita and H. Kitano. Development of an autonomous quadruped robot for robot entertainment. *Autonomous Robots*, 5(1):7–18, 1998.
13. K. George. *Malayalam Grammar and Reader.* National Book Stall, Kottayam, 1971.
14. A. E. Goldberg. *A Construction Grammar Approach to Argument Structure.* University of Chicago Press, Chicago, 1995.
15. W. Haeseryn, K. Romijn, G. Geerts, J. de Rooij, and M. van den Toorn, editors. *Algemene Nederlandse Spraakkunst.* Martinus Nijhoff / Wolters Plantyn, Groningen/Deurne, second, revised edition, 1997.
16. M. A. Halliday. *An Introduction to Functional Grammar.* John Benjamins, Amsterdam, 1994.
17. M. Hickmann and S. Robert, editors. *Space in Languages.* Typological Studies in Language 66. John Benjamins, Amsterdam, 2006.
18. P. Hopper. Emergent grammar. *Berkely Linguistics Conference (BLS)*, 13:139–157, 1987.
19. T. Iachini and R. H. Logie. The role of perspective in locating position in a real-world, unfamiliar environment. *Applied Cognitive Psychology*, 17(6):715–732, 2003.
20. I. Jung. *Grammatik des Paez. Ein Abriss. PhD Thesis.* University of Osnabrück, Osnabrück, 1989.
21. M. Jüngel, J. Hoffmann, and M. Lötzsch. A real-time auto-adjusting vision system for robotic soccer. In D. Polani, B. Browning, and A. Bonarini, editors, *RoboCup 2003: Robot Soccer World Cup VII*, volume 3020 of *Lecture Notes in Artificial Intelligence*, pages 214–225, Padova, Italy, 2004. Springer.
22. P. Kay and C. J. Fillmore. Grammatical constructions and linguistic generalizations: The what's x doing y? construction. *Language*, 75:1–33, 1999.
23. S. Kuno. *The Structure of the Japanese Language.* MIT Press, Cambridge, MA, 1973.
24. B. Landau and R. Jackendoff. 'what' and 'where' in spatial language and spatial cognition. *Behavioural and Brain Sciences*, 16:217–238, 1993.
25. R. W. Langacker. *Foundations of Cognitive Grammar. Volume 1.* Stanford University Press, Stanford, 1987.
26. S. C. Levinson. *Space in Language and Cognition.* Language, Culture and Cognition 5. Cambridge University Press, Cambridge, 2003.
27. S. C. Levinson and D. Wilkins, editors. *Grammars of Space. Explorations in Cognitive Diversity.* Language, Culture and Cognition. Cambridge University Press, Cambridge, 2006.
28. F. Lichtenberk. *A Grammar of Manam.* Oceanic Linguistics Special Publications 18. University of Hawaii Press, Honolulu, 1983.

29. M. Loetzsch, M. Risler, and M. Jüngel. XABSL - a pragmatic approach to behavior engineering. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2006)*, pages 5124–5129, Beijing, October 2006.
30. C. Mair. Corpus linguistics and grammaticalization theory: Beyond statistics and frequency?corpus linguistics and grammaticalization theory. In H. Lindquist and C. Mair, editors, *Corpus Approaches to Grammaticalization in English*, pages 121–150. John Benjamins, Amsterdam, 2004.
31. B. Mauck and G. Dehnhardt. Mental rotation in a california sea lion (zalophus californianus). *Journal of Experimental Biology*, 200(9):1309–1316, 1997.
32. J. Moshinsky. *A Grammar of Southeastern Pomo.* University of California Press, Los Angeles, 1974.
33. M. Mous. *A Grammar ofd Iraqw. PhD Thesis.* Rijksuniversiteit Leiden, Leiden, 1992.
34. J. Nuyts. *Aspects of a Cognitive-Pragmatic Theory of Language. On Cognition, Functionalism, and Grammar.* John Benjamins, Amsterdam, 1992.
35. J. Nuyts. Brother in arms? on the relations between cognitive and functional linguistics. In F. Ruiz de Mendoza Ibanez and M. Pena Cervel, editors, *Cognitive Linguistics: Internal Dynamics and Interdisciplinary Interaction*, pages 69–100. Mouton De Gruyter, Berlin, 2005.
36. S. Pinker and P. Bloom. Natural language and natural selection. *Behavioral and Brain Sciences*, 13(4):707–784, 1990.
37. T. Röfer, R. Brunn, I. Dahm, M. Hebbel, J. Hoffmann, M. Jüngel, T. Laue, M. Lötzsch, W. Nistico, and M. Spranger. Germanteam 2004: The german national robocup team. In D. Nardi, M. Riedmiller, C. Sammut, and J. Santos-Victor, editors, *RoboCup 2004: Robot Soccer World Cup VIII*, volume 3276 of *Lecture Notes in Artificial Intelligence*, Lisbon, Portugal, 2005. Springer.
38. R. Shepard and J. Metzler. Mental rotation of three-dimensional objects. *Science*, 3(171):701–703, 1971.
39. L. Steels. Perceptually grounded meaning creation. In M. Tokoro, editor, *Proceedings of the International Conference on Multi-Agent Systems*, pages 338–344. The MIT Press, 1996.
40. L. Steels. How to do experiments in artificial language evolution and why. In A. Cangelosi, A. Smith, and K. Smith, editors, *Proceedings of the 6th International Conference on the Evolution of Language*, London, 2006. World Scientific Publishing.
41. L. Steels and J. De Beule. Unify and merge in fluid construction grammar. In P. Vogt, Y. Sugita, E. Tuci, and C. Nehaniv, editors, *Symbol Grounding and Beyond: Proceedings of the Third International Workshop on the Emergence and Evolution of Linguistic Communication, EELC 2006*, volume 4211 of *Lecture Notes in Artificial Intelligence*, pages 197–223, Rome, Italy, September 2006. Springer Verlag.
42. L. Steels, F. Kaplan, A. McIntyre, and J. Van Looveren. Crucial factors in the origins of word-meaning. In A. Wray, editor, *The Transition to Language.* Oxford University Press, Oxford, UK, 2002.
43. L. Steels and M. Loetzsch. Perspective alignment in spatial language. In K. R. Coventry, T. Tenbrink, and J. A. Bateman, editors, *Spatial Language and Dialogue.* Oxford University Press, 2007. to appear.
44. A. Stefanowitsch and S. T. Gries. Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics*, 2(8):209–243, 2003.

45. L. Talmy. *Toward a Cognitive Semantics, Concept Structuring Systems*, volume 1. MIT Press, Cambridge, Mass, 2000.
46. L. Talmy. *Toward a Cognitive Semantics, Typology and Process in Concept Structuring*, volume 2. MIT Press, Cambridge, Mass, 2000.
47. L. Talmy. The fundamental system of spatial schemas in language. In B. Hampe, editor, *From Perception to Meaning: Image Schemas in Cognitive Linguistics*, Cognitive Linguistics Research, pages 37–47. Mouton De Gruyter, Berlin, 2006.
48. M. Tomasello. Joint attention as social cognition. In C. Moore and P. J. Dunham, editors, *Joint Attention: Its Origins and Role in Development*. Lawrence Erlbaum Associates, Hillsdale, NJ, 1995.
49. E. C. Traugott and B. Heine, editors. *Approaches to Grammaticalization*, volume 1 of *Typological Studies in Language 19*. John Benjamins, Amsterdam, 1991.
50. H. Werner. *Die Ketische Sprache*. Tunguso Sibirica Band 3. Harrasowitz Verlag, Wiesbaden, 1997.
51. J. Zacks, B. Rypma, J. Gabrieli, B. Tversky, and G. H. Glover. Imagined transformations of bodies: An fMRI investigation. *Neuropsychologia*, 37(9):1029–40, August 1999.
52. F. Zewen. *Introduction à la Langue des Îles Marquises. Le Parler de Nukuhiva.* University of Hamburg, Hamburg, 1987.