Argument linking and spatial cognition

Alistair Knott (alik@cs.otago.ac.nz)
Dept of Computer Science, University of Otago, New Zealand

Abstract

This paper presents evidence that certain syntactic phenomena relating to the GB model of argument structure have their origins in a non-linguistic model of spatial cognition.

1 Introduction

The general hypothesis proposed in this paper is that an utterance in natural language carries information not directly about the world, but rather about a cognitive process occurring in some agent (the speaker, or perhaps the protagonist in a close-perspective narrative). An indicative sentence, for instance, tells us not directly about a situation in the world, but about the perceptual process by which an agent obtains information about this situation—a process which involves directions of attention, and spatial cognitive representations in long-term and short-term memory. The syntactic and semantic structure of a sentence can therefore be expected to contain reflexes of the underlying perceptual process. This hypothesis places the paper within a broad tradition in linguistics, encompassing the school of 'cognitive linguistics' on the one hand (see e.g. Langacker, 1987) and the work of Jackendoff and colleagues on the other (see e.g. Jackendoff, 1983).

The hypothesis under investigation can be studied in its most concrete form by concentrating on the structure of sentences which express simple spatiotemporal eventualities, apprehendable directly by visual perception. Vision is the perceptual modality about which most is known, and sentences expressing visually perceivable facts can certainly furnish us with a sufficiently large range of interesting linguistic phenomena to study. If the hypothesis can be supported for concrete sentences of this kind, an extension to more abstract sentences can be envisaged, perhaps drawing on the idea that space is used as a device for structuring abstract domains (see e.g. Lakoff and Johnson, 1980). I will begin in Section 2 by describing the linguistic phenomena I will concentrate on, which focus on the roles of the subject and verb phrase in the syntax and semantics of existential sentences. In Section 3, I outline a model of the spatial representations mediating these sentences, motivated from recent work in visual attention. In Section 4, I argue that this model can play a role in the explanation of the linguistic data.

2 Some linguistic data

I will focus on the following set of sentences, which express concrete spatiotemporal eventualities, and also illustrate some much-studied linguistic phenomena.

(1) [A dog]$_F$ is in the room.

(2) There is [a dog]$_F$ in the room.

(3) A dog has [a collar]$_F$.

(4) *There is [a dog]$_F$ hairy.

(5) The dog is [in the room]$_F$.

(6) The dog has [a collar]$_F$.

(7) *There is [the dog]$_F$ in the room.

A linguistic model of these sentences has to account for several things. Firstly, Examples 1 and 2 show that an existential sentence introducing a new individual into a discourse can be expressed in two ways, either with an indefinite subject NP, or the expletive subject NP *there.* A standard explanation in GB (see e.g. Stowell, 1978; Burzio, 1986) is that these two structures are associated with the same DS structure. *Be* is analysed as a raising verb with no external argument, subcategorising for a small clause whose subject *a dog* can optionally move to the empty external argument position.

Secondly, these sentences illustrate a difference between **stage-level** predicates denoting transitory properties like 'being in the room' and **individual-level** predicates denoting permanent properties like 'having a collar' or 'being hairy' (c.f. Carlson, 1977). When an indefinite NP is subject of a stage-level predicate, as in Example 1, it is interpreted existentially, introducing a new individual into the discourse; when an indefinite NP is subject of an individual-level predicate, as in Example 3, the sentence is interpreted as a generic proposition. This data is commonly analysed by making reference to Diesing's (1992) 'mapping hypothesis', which states a structural correspondence between a sentence's DS representation and its LF representation (which consists of a tripartite structure: a **quantifier**, a **restrictor set** and a **nuclear scope set**). The hypothesis suggests that material from the [Spec,IP] (Specifier-of-IP) at DS is mapped to the restrictor set and material from the DS VP is mapped to the scope set. Unselective quantification ensures that all variables appearing in the restrictor set are bound by the quantifier, and any occurrences of these variables in the scope set are likewise bound. Any remaining unbound variables in the scope set are taken to be bound by existential quantification. The suggestion is then that the subjects of stage-level predicates are base-generated within the VP, while those of individual-level predicates are base-generated outside the VP (at [Spec, IP]). An indefinite NP contributes an unbound variable: if this indefinite is the SS subject of an individual-level predicate, as in Example 3, it has been base-generated in this position, and thus contributes the variable to the restrictor set, where it is bound by a quantifier (in Example 3 a non-overt generic one) and receives a generic reading. If the indefinite NP is the subject of a stage-level predicate, as in Example 1, it has moved to this position from within the VP, and can 'lower' back to the VP at LF. In this scenario it contributes to the nuclear scope set, where it is bound by existential closure, thus receiving an existential reading.

Thirdly, Examples 2 and 4 illustrate the fact that stage-level predicates can, and individual-level predicates cannot, appear in *there*-sentences. An explanation of this fact is likely to make reference to the VP-internal origins of the subjects of stage-level predicates. (It should be noted that not all stage-level predicates can appear in *there*-sentences; however, we will only be concerned with those which can.) Finally, while both stage-level and individual-level predicates can take a definite subject NP, as in Examples 5 and 6, Example 7 shows that a definite NP (or more generally, a strong NP) cannot appear in a *there*-construction.

3 A deictic model of visual attention

Examples 1–7 are based on eventualities which can be directly apprehended visually: noticing the existence of a dog in a room, and noting a physical property (having a collar, being hairy) of an observed individual. I will now outline a high-level model of the visual processes that underlie the apprehension of such events, supported by experimental work in visual attention.

It is well known that vision is an active process: we perceive the world by attending to a succession of different points in the visual array, rather than by attending uniformly to all points in the array (see e.g. Ballard, 1991). The active nature of vision raises the question of how the transitory **retina-centred representation** of the environment (RCR) is integrated into more stable and enduring representations, and what these representations are. There is evidence for at least three constructed representations; I will describe these below in turn.

An **object-centred representation** (OCR) is an agent's internal cognitive representation of an object, given in its own coordinate system. The process of visual object recognition is by definition the process of mapping a 2D retinal projection onto a suitable OCR; the interesting question is, of course, what type of representation OCRs provide, and this question is still the focus for much debate in vision research. The two main candidates are 3D volumetric representations (see e.g. Biederman, 1985) and 'multiple views' representations (Ullman and Basri, 1991). I will adopt the former model here, mainly because of its conceptual simplicity; it allows us to describe the process of fitting an object model to a foveated retinal stimulus as the process of finding a suitable 3D transformation (including a rotation and a magnification as principal components) to project a model onto a 2D retinal stimulus. I will refer to volumetric object models as object **templates**.

An **environment-centred representation** (ECR) is a representation of the agent's environment in a coordinate system centred within that environment, commonly known as a cognitive map. It provides a representation of the location (and possibly orientation) of objects in this environment, including a representation of the agent's own location. There is good evidence that such a representation is provided by the human and infrahuman hippocampus (O'Keefe and Nadel, 1978). We will assume that an object in the ECR is specified by an association between a unit denoting a particular point in the coordinate system and the agent's internal model of this object, such that activation of the model causes activation of this unit, and vice versa. A central function of the ECR is to facilitate reattention to known objects within a scene. If the representation specifies the agent's gaze direction as well as her location (for which there is also good evidence) then it can be used to calculate the change in gaze direction needed to foveate any of the other objects which feature in the representation.

A final frame of reference is termed the

index-centred representation. This is a representation of a region of the agent's local environment, given in terms of a coordinate system centred on one of the objects in this region, which we will term the **primary index** object. It differs from the OCR in that it represents not only the geometry of the index object using its own coordinate system, but also uses the same coordinate system to represent the position and relative orientation of salient objects in the environment of the index object. Perceptual evidence for the ICR comes primarily from a perceptual disorder called object-centred neglect, associated with damage to the left or right parietal cortex (Behrmann and Tipper, 1999). Patients with this condition have systematic difficulties noticing stimuli on the contralesional side of an object they are attending to; for instance, patients asked to count occurrences of the letter A on a TV screen might systematically ignore the letters on the left side of the screen. Neglect has to be described using object-centred coordinates rather than retina-centred ones, because it persists even in the presence of eye movements which place the target object on different parts of the retina at different moments. Even more strikingly, neglect must still be described in object-centred terms even if the patients observe the target rotating through 180°; if the intitial deficit was contralesional, after such a rotation, it beomes ipsilesional. The stability of the ICR over saccades makes it a good representation to use for the planning of sequences of saccades; 'virtual saccades' to points on the ICR can be converted into actual saccades by a simple process of adding them to V_{fi}. Indeed, neuropsychological evidence for such virtual saccades in the infrahuman visual system has recently been found in single-cell recordings (Olson and Gettner, 1995).

What is the purpose of the representation provided by the ICR? We will assume that it is primarily used for recognising configurations of an articulated object (e.g. postures of a human agent). This task would certainly be simplified if we could assume a stabilised representation of the object. If we take the ICR to be centred on a person's torso, for instance, then the configuration of her limbs given by the ICR will be reasonably similar for any given posture, regardless of her position and orientation in relation to the viewer. The argument can be extended to configurations involving additional objects, such as the configuration of a person holding an object, or a dog having a collar, and also to changes in configuration over time, i.e. actions. Accordingly, we will postulate that the ICR is the representation which our internal templates for spatial verbs 'look at' when an action or configuration of an attended-to agent or object is being categorised. The stability of the ICR over saccades makes it especially suitable for this role. Spatial verbs can have complements, corresponding to secondary objects to be attended to, and there must be a way of planning saccades to such objects without disrupting the ongoing representation of the index object and its configuration.

3.1 Integrating information between coordinate systems

A summary of the model of spatial cognition is given in Figure 1. Links between representations are bidirectional; e.g. information in the OCR can be conveyed to the ICR and vice versa. Note that there are two links between the RCR and the ICR: one link goes via object recognition (i.e. via the OCR) and the other is direct. Evidence for the direct link comes for instance from the classic experiments of Johansson (1973), which suggest that the action of an animate object can be recognised prior to, or at least independently of, recognition of the object itself.

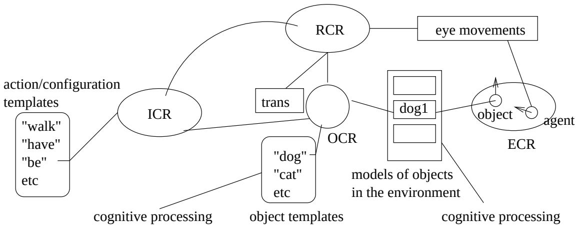
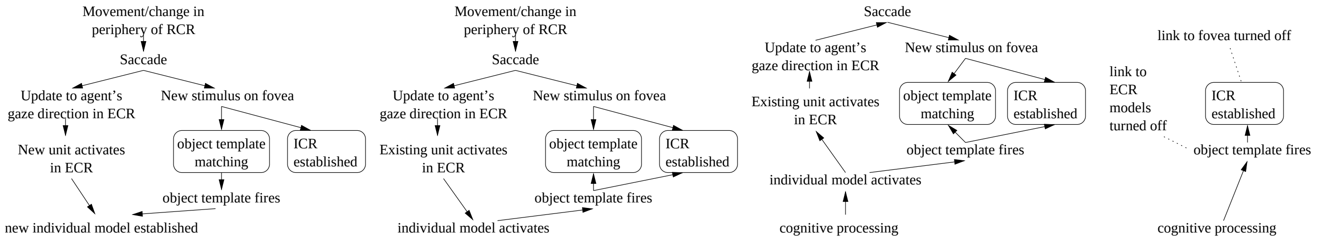


Figure 1: Links between the coordinate systems

Representations in each coordinate system are **deictic**, in the sense of Ballard *et al.* (1997), in that they make implicit reference to other representations being computed in real time in the system. For instance, activation of an object template needs to be interpreted in the light of constantly updated information about the agent's gaze direction given in the ECR. In general, changes in one representation have to be communicated to other representations to maintain consistency in the system. The directions of communication between representations depend on where the initial change occurs, as shown in Figure 2, and as described below.

Bottom-up attention to a new object
In the scenario in Figure 2(i), an action of attention begins with a saccade triggered bottom-up by a strong stimulus in the retinal periphery, caused by an object not yet represented in


Movement/change in periphery of RCR
Saccade
Update to agent's gaze direction in ECR
New stimulus on fovea
New unit activates in ECR
object template matching
ICR established
object template fires
new individual model established
Movement/change in periphery of RCR
Saccade
Update to agent's gaze direction in ECR
New stimulus on fovea
Existing unit activates in ECR
object template matching
ICR established
object template fires
individual model activates
Saccade
Update to agent's gaze direction in ECR
New stimulus on fovea
Existing unit activates in ECR
object template matching
ICR established
object template fires
individual model activates
cognitive processing
link to fovea turned off
link to ECR models turned off
ICR established
object template fires
cognitive processing

(i) Bottom−up attention to a new object

(ii) Bottom−up attention to a known object

(iii) Top−down attention to a known object

(iv) Top−down attention to an object template

Figure 2: Patterns of communication between coordinate systems

the observer's ECR. The saccade results in two things: a newly active ECR unit, and (when object categorisation is completed) a new template and transformation associated with the fovea and with the index of the ICR. The result of this process is the linking of this template with the individual associated with the new ECR unit, to facilitate subsequent re-attention to the object it represents. This scenario will be argued to underlie Examples 1 and 2.

Bottom-up attention to a known object In Figure 2(ii), a saccade triggered bottom-up by a peripheral stimulus causes the activation of an ECR unit already associated with an object model. In this case, the template linked to this model can be associated immediately with the ICR and the retina-based object recognition machinery, to bring a top-down component to the process of recategorisation. Here, the causality between ECR model and ICR template runs in the opposite direction from Scenario (i), with the model informing the template. This scenario will be argued to underlie Example 5.

Top-down attention to a known object In Figure 2(iii), the action of attention is initiated by a top-down mechanism: the activation of an internal representation of an object in the current scene. This activation triggers two things simultaneously: one is a saccade (as a result of the link between the model of the object and a unit in the ECR), and the other is the top-down activation of an appropriate template, to facilitate recategorisation, as above. This scenario differs from (i) and (ii), in that the saccade is an effect, rather than the primary cause. It will be argued to underlie Example 6.

Top-down attention to an object template In Figure 2(iv), an action of attention is initiated by the top-down activation of a first-order template with no link either to the retina or to a model associated with an ECR unit. We suggest that actions of this kind are associated with our ability to talk and reason generically about kinds of objects: the suggestion is basically that the agent is able to attend to her internal model of an object using the same architecture that is involved in attention to concrete objects in the world. This possibility can certainly be accomodated within the present model, and indeed within any model which features a level of representation integrating top-down information due to a template with bottom-up information deriving from the retina. This scenario will be argued to underlie the generic Example 3.

3.2 Summary

There are two interesting points to note about the role of the ICR in the scenarios just described. Firstly, an action of attention can be seen as comprising two distinct stages: an **initial phase**, in which a pattern of stimulation is established on the ICR, and a **subsequent phase** driven by the ICR, in which action templates are categorised. (The latter phase optionally involves the initiation of further saccades which are planned in ICR-centred coordinates; these are not shown in the diagrams in Figure 2.) Secondly, note that the point at which an object template is associated with the primary index object can vary: in Scenario (i) it occurs in parallel with processing in the ICR, while in (ii)–(iv) it occurs during the initial phase, before ICR-based processing begins. The intuition behind Scenario (i) is that an object can be fixated and tracked, and the action/configuration in which it is involved can begin to be processed, before the object is actually categorised; thus, the model allows us to talk about the influence of an 'uncategorised' retinal stimulus on processing in the ICR.

4 Syntactic reflexes of the attentional model

The central claim of this paper is that the concepts of subject and verb phrase, as formulated within GB theory, can be linked to descriptions of processing within the model of spatial cognition just given, at least for sentences describing simple spatial eventualities.

I will assume, following the discussion in Section 1, that the semantics of a sentence describing a spatiotemporal event is a trace of the perceptual processing which led to the event being apprehended by an observer. Given our model of visual attention, the trace will involve a description of a sequence of events, similar to those given in Scenarios (i)–(iv) outlined above. Recall that these scenarios all involve a distinction between an initial phase of establishing a representation in the ICR, and a subsequent phase of processing within the ICR. Recall also that the process of associating a template with the ICR's primary index object can occur as part of the initial phase (in Scenarios (ii)–(iv)) or as part of the subsequent phrase (in Scenario (i)). On the basis of these assumptions, I propose that the syntactic structure of a sentence describing a concrete spatiotemporal event can be given a spatial characterisation as follows.

Definition 1. At DS, [Spec,IP] contributes information about processing during the initial phase of the action of attention which underlies the sentence. DS VP contributes information about processing during the subsequent phase of the action of attention.

Definition 2. An argument position at DS (e.g. [Spec,IP]) contributes information about a process of object categorisation culminating in the association of an object template with the OCR. An empty argument position at DS identifies a point in this process where an object has been attended to, but is not yet associated with an object template, while a full NP signifies the point at which an object template is associated with the OCR.

Definition 3. Stage-level predicates (at least those which can appear in *there*-contexts) denote events or properties of objects which are salient enough to initiate bottom-up actions of attention to points in the retinal periphery (Scenarios (i) or (ii)). Individual-level predicates denote properties which can only be registered top-down, by directions of attention to known objects (Scenarios (iii) or (iv)).

Definition 4. All NPs describe the activation of an object template. The difference between definite and indefinite NPs relates to the cause of the template's activation. A definite NP describes a scenario where an object template is activated by the activation of an individual model associated with a unit in the ECR, as in Scenarios (ii) or (iii). An indefinite NP denotes a scenario where an object template becomes active for some other reason: either due to bottom-up stimuli from the retina, as in Scenario (i), or due to direct top-down activation as in the generic Scenario (iv).

Definitions 1–3 can be thought of as a reformulation of Diesing's idea that the surface subject of a stage-level predicate originates within VP at DS while that of an individual-level predicate originates in [Spec,IP]. By Definition 3, a stage-level predicate can initiate an action of attention bottom-up to a point on the retina which is not yet associated with an object template: Scenario (i) in our model. By Definitions 1 and 2, a sentence describing this action of attention will have an empty [Spec,IP] at DS, since the attended-to object is not associated with a template during the initial phase. This is indeed what we see in Examples 1 and 2. By Definition 3, a sentence with an individual-level predicate must describe a top-down action of attention, so the template of the index object is known during the initial phase: hence it has to have a non-empty [Spec,IP] at DS.

Definition 4, taken together with Definitions 1–3, generates the appropriate predictions about the behaviour of definite and indefinite subject NPs in the context of stage-level and individual-level predicates and *there*-contexts. In our model, a subject NP base-generated within the VP is only possible in the case of a bottom-up action of attention to an unknown object, i.e. in Scenario (i). In this scenario, the NP describes the activation of an object template by stimuli on the retina; by Definition 4, such an NP has to be indefinite. A subject NP base-generated in [Spec,IP] can either be definite or indefinite. If it is definite, by Definition 4 there are two possibilities: Scenario (ii), which is initiated bottom-up (and hence must be stage-level) and Scenario (iii), which is initiated top-down. Examples 5 and 6 show that both stage- and individual-level predicates can have definite subjects base-generated in [Spec,IP]. Finally, if the NP base-generated in [Spec,IP] is indefinite, there is only one possibility. By Definition 4, an indefinite NP can only occur in Scenarios (i) or (iv). We have already seen that in Scenario (i), the subject NP is base-generated within the VP; this only leaves Scenario (iv). We thus have an explanation of how it is that a sentence with an individual-level predicate and an indefinite subject NP has to be interpreted as a generic.

We still have to provide an account of what it means to say that a subject NP originating within the VP at DS can 'move' to [Spec,IP] at SS. I suggest that this movement is a reflection of the fact that it is not possible to formulate a trace of an action of perception as a linear sequence of events, due to the deictic nature of the representations involved. The difficulty is in characterising the object attended to in the initial phase of the action of attention. If we refer to it by its template, we lose the portion of the account which tells us that it is uncategorised at the point when it establishes the ICR. If we refer to it using an expletive, we capture this component of the process, but lose the component during which it is associated with a template. To solve the problem, two references are made to this object in the DS structure of the sentence: one VP-external, giving its role as the initiator of the ICR, and one VP-internal, describing its categorisation using an object template. Given that these are both references to the same object, albeit at different points in time, we can construe the SS difference between Example 1 (with no NP raising) and Example 2 (with NP raising) as simply a matter of whether the action of attention is described 'progressively', as a temporal sequence of states, or 'retrospectively', using the finished states of all the relevant components of the system.

5 Conclusions

The aim of this paper has been to flesh out a contemporary GB treatment of existentials, generics and definiteness by grounding certain syntactic structures in an independently-motivated psychological account of visual attention. The evidence adduced here is only very preliminary, but it is nonetheless quite promising. Syntactic structures are notoriously resistant to semantic characterisation, but there does seem some potential in associating DS subject and DS VP with the initial and subsequent phases of a perceptual action. One particular advantage of the proposed model is that the semantic representations of existential sentences and of generic sentences are closely tied to the structural account. The syntactic model of existential sentences describes a mechanism by which objects not yet known to an observer can attract attention to themselves, which seems a good basis for a semantic account of a sentence introducing a new entity into the discourse. And the syntactic model of generic sentences contains within it a plausible semantic account, in which generics are predicates about an agent's internal models of objects.

References

Ballard, D. (1991). Animate vision. *Artificial Intelligence*, **48**, 57–86.

Ballard, D., Hayhoe, M., Pook, P., and Rao, R. (1997). Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences*, **20**(4), 723–767.

Behrmann, M. and Tipper, S. (1999). Attention accesses multiple reference frames: Evidence from visual neglect. *Journal of Experimental Psychology*, **25**(1), 83–101.

Biederman, I. (1985). Human image understanding: Recent research and a theory. *Computer Vision, Graphics and Image Processing*, **32**, 29–73.

Burzio, L. (1986). *Italian Syntax. A Government-Binding Approach*. Kluwer, Hingham, MA.

Carlson, G. (1977). *Reference to Kinds in English*. Ph.D. thesis, University of Massachusetts at Amherst.

Diesing, M. (1992). *Indefinites*. MIT Press, Cambridge, MA.

Itti, L. and Koch, C. (in press). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*.

Jackendoff, R. (1983). *Semantics and cognition*. MIT Press, Cambridge, MA.

Johansson, G. (1973). Visual perception of biological motion, and a model for its analysis. *Visual Perception and Psychophysics*, **14**, 201–211.

Lakoff, G. and Johnson, M. (1980). *Metaphors we live by*. University of Chicago Press, Chicago and London.

Langacker, R. (1987). *Foundations of cognitive grammar I: Theoretical prerequisites*. Stanford University Press.

Milsark, G. (1974). *Existential sentences in English*. Ph.D. thesis, MIT, Cambridge, MA.

O'Keefe, J. and Nadel, L. (1978). *The hippocampus as a cognitive map*. Clarendon Press, Oxford.

Olson, C. and Gettner, S. (1995). Object-centred direction sensitivity in the Macaque supplementary eye field. *Science*, **269**, 985–988.

Rao, R. and Ballard, D. (1996). A computational model of spatial representations that explains object-centred neglect in parietal patients. In *Proceedings of Computational Neuroscience*.

Stowell, T. (1978). What was there before there was there. In *Proceedings of the Chicago Linguistic Society*, volume 14, pages 458–471.

Ullman, S. and Basri, R. (1991). Recognition by linear combinations of models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **13**(10), 992–1006.

Zamparelli, R. (2000). *Layers in the Determiner Phrase*. Garland, New York.