

# Objects and Nouns: Ontologies and Relations

Francesco-Alessio Ursini & Aijun Huang

English Department, Stockholm University & English Department, Soochow University  
Sodra Husen E, E870, Frescativagen, Stockholm & Suzhou City, 10 Azusa Street, Shanghai  
[francesco.ursini@english.su.se](mailto:francesco.ursini@english.su.se) & [ajhuang@suda.edu.cn](mailto:ajhuang@suda.edu.cn)

## 1. The problem

A well-known problem in cognitive science is the definition of empirically motivated ontologies that represent the building blocks and relations underpinning our cognitive domains (Smith, 1995). Ontologically-driven cognitive approaches have suggested that cross-domain categories may include internalized representations of *objects*, *events* and *places*, as such basic blocks (Metzinger & Gallese, 2003). Formal models in AI often involve attribute-driven ontologies and taxonomies (Olszewka, 2011). Nevertheless, two questions seem still outstanding. A first is whether formal ontological models can be grounded on empirically based cognitive analyses. A second is whether the relations amongst ontological units so defined can also model the relations amongst different cognitive domains.

The goal of this paper is to answer these two questions by concentrating on two cognitive domains: vision and language. We opt to focus on two categories: *visual objects* (henceforth: VOs) and *noun phrases* (henceforth: NPs). Both categories are well-studied, but a theory on how these categories are related is still outstanding. Hence, we offer an account of the ontological structures underpinning these categories and their relation(s), based on the notion of *informorphism*. We then show how this account can help us in offering a solution to three problems regarding the relation between VOs and NPs.

## 2. An Ontology of VOs

Classical analyses of object recognition observed that formal definitions of “objects” are far from trivial (Marr, 1982: 222). Current proposals, however, suggest that VOs are any visual *tokens* that can be individuated via some discriminating property or *type*, and can be *loci* of attention (Scholl, 2007). Four parallel sub-systems seem to govern how this processes occur, and individuate VOs.

A first system seems to identify VOs via their spatio-temporal trajectories. VOs may be perceived as moving independently or not, hence having or lacking agency. Shapeless, colour-less dots can still be individuated as VOs, insofar as they follow a trajectory (Gao et al., 2009). Thus, *motion* acts as a type of individuating principle for VO tokens. A second system can instead individuate objects via the attributes or *textons* they instantiate (Zhu et al., 2005). Textons include features such as the material, colour or texture that objects are made of. Thus, a shapeless amount of water can count as a token VO of the *texton* type, *material* sub-type.

A third system seems instead to individuate VOs by their *shape*. Models such as RBC (Biederman, 1987) suggest that basic shapes or *geons* can be defined as the combination of basic geometrical properties (e.g. collinearity). Geons can in turn be combined to form complex VOs, belonging to various shape (sub-)types (e.g. human figures, cars). A fourth system seems to act in parallel to these three systems, and seems to establish a

“counting” structure for VOs via indexing mechanisms (*Fingers of INSTantiation* or FINSTs: Scholl, 2007). For instance, a VO token corresponding to three pears can instantiate these pears as objects counted individually, or as a multiplicity, or even as a single group type (Scholl, 2007). Quantity-specific types of VOs can also be individuated (e.g. “three” or “most” pears), when verbal instructions introduce such fine-grained individuation principles (Pietroski et al., 2009). Thus, a VO can instantiate different sub-types of a general *quantity* type.

In order to model the types that these systems seem to individuate, we propose to use a relational Boolean algebra  $VO = \langle O, O^*, \sqsubseteq \rangle$  as an information structure (Barwise & Seligman, 1997). This structure includes a set of tokens  $O$ , a set of basic types  $O' = \{motion, texton, shape, quantity\}$ , and a power-set  $O^*$  of complex types, on which the operations *meet* ( $\sqcap$ ) and *join* ( $\sqcup$ ) are defined. We represent sub-types as “values” of types, e.g. *blue:texton* for a colour texton. Sets of possible values for a type are represented via the *join* operator: *red*  $\sqcup$  *blue* is a minimal set of *colour:texton* type values. Complex types are represented via the *meet* operation, e.g. *shape*  $\sqcap$  *texton*. A *part-of* relation between types can be defined, e.g.  $a \sqsubseteq a \sqcap b$  ( $a \sqsubseteq a \sqcup b$ ), which reads: the type  $a$  is part of the complex type  $a \sqcap b$  ( $a \sqcup b$ ). When a token is part of a type, we say that it *instantiates* a type, and represent it as  $x \sqsubseteq type$ . In our example, the relation  $x \sqsubseteq shape:pear \sqcap \beta:quantity$  holds: a VO token  $x$  instantiates a type of object with the shape of a pear, and a cardinality (sub-)type “three”.

## 3. An Ontology of NPs

We focus on NPs in subject position in predicative constructions, and individuate semantic types of this near-universal morpho-syntactic category (Baker, 2003). Our key examples are in (1)-(6), below:

- (1) The crow is flying over the cloud
- (2) The red ball rolled below the table
- (3) Water was pouring through the hole
- (4) The crows are flying over the cloud
- (5) The group of crows was flying over the cloud
- (6) At least three crows are flying over the cloud

Consider (1)-(3), first. The (subject) definite NPs *the crow* and *the red ball* in (1)-(2) alternate with respect to a sense dimension of *animacy*: whether a noun denotes an independent agent (*the crow*) or not (*the ball*) (Yamamoto 1999). Both NPs belong to a *sortal* super-type, as their senses individuate referents via their shape and/or attributes (e.g. colour: Pelletier, 2009). Both NPs also belong to the singular (quantity) type: they denote single, “atomic” crows and (red) balls, respectively. Instead, the sense of *water* in (3) denotes a morphologically singular but semantically *mass* term, a term that individuates a discourse referent by its *properties* (Pelletier 2009).

Examples (4)-(6) suggest that English also has a rich

system to express quantity types, beside the *singular* type. In (4)-(5), the definite NPs *the crows* and *the group of crows* respectively denote a plurality of crows (i.e. sum of individual crows) and their corresponding group (Landman 2000). In (6), the quantified NP *at least three crows* individuates a precise quantity of crows in discourse. Many other examples of quantified NPs (e.g. *all the crows*) could be offered, but the examples should already illustrate the point (Peters & Westerståhl, 2006).

The examples (1)-(6) do not aim to be exhaustive, but seem to support a formal analysis of these types and their relations via a Boolean algebra, too. Thus, we suggest that we have a structure  $NP = \langle N, N^*, \subseteq \rangle$ , with  $N$  a set of tokens,  $N^*$  a basic set of NP types defined as  $N^* = \{animacy, property, sort, quantifier\}$ , from which a power-set of types  $N^*$  is generated. These types can also include sub-types (e.g. *plu:quant* for plural NPs such as *crows*), and join and meet types: *crow*  $\sqcup$  *ball* are a minimal set of values for *shape*, *red*  $\sqcap$  *ball* is a complex type. An NP can then be said to instantiate a certain type, if it belongs to it: we have  $(the\ red\ ball) \sqsubseteq red:property \sqcap ball:sort$ .

#### 4. A Proposal: an Infomorphic Approach

We can now offer our answer to our two questions, which is defined as follows. An infomorphism consists of a pair of mappings between information structures, in this case our relational Boolean algebras (Barwise & Seligman, 1997: ch.1-3). Given our structures  $VO = \langle O, O^*, \subseteq \rangle$  and  $NP = \langle N, N^*, \subseteq \rangle$ , we define a pair of functions  $\langle f, g \rangle$  such that  $VO = \langle f(N), O^*, \subseteq \rangle$  if and only if  $NP = \langle g(O), N^*, \subseteq \rangle$ . In words, VO tokens can instantiate VO types, which are mapped onto NP types, in turn instantiated by NP tokens. Tokens of both domains, via this pair of mappings, can be thus connected:  $f$  maps VO tokens to NP tokens,  $g$  outlines the reverse mapping. In words, “objects” can be identified via the NPs that describe them and *vice versa*. Since we are now able to establish a mapping between these different cognitive/ontological structures, we are able to solve at least three problems regarding this relation.

First, we can account certain problems of *reference* (Abbott, 2010): for instance, how different NPs types can refer to the same VO in an extra-linguistic context. Suppose that we have a scenario in which we see several crows flying over a cloud, and that we generate a VO token  $a$  for these crows. The instantiation relation  $a \sqsubseteq crow:shape \sqcap quantity$  holds for two *quantity* values: *plural* or *group*. We represent this relation as  $a \sqsubseteq ((crow:sh \sqcap (pl \sqcup gr):quant))$ . In words, we can focus on these crows as a multiplicity or as a group. Suppose, then, that we either use (4) or (5) to describe this scenario. The two NPs *the crows* and *the group of crows* are tokens that respectively belong to the types  $(crow:sort \sqcap plur:quant)$  and  $(crow:sort \sqcap group:quant)$ . Thus, a pair of functions  $f$  and  $g$  can map the same VO token with two distinct NP tokens: *the crows* and *the group of crows*. As a result, distinct NPs can refer to the same VO token.

Second, cross- and intra-linguistic *lexical variation* finds a related analysis. For instance, if we take some amount of money we have in our pocket, we can use the NPs *coins* or *change* to refer to them (Pelletier 2009). While *coins* is a countable and plural NP, *change* is uncountable (mass) NP (cf. *how many coins* vs. *how much change*). Both can refer to the same (VO) token, but so does Mandarin *Jinqián*. However, this Mandarin NP

token is lexically ambiguous: it can instantiate both countable (plural) and uncountable values of the *quantifier* type. The type relations that hold for these NPs are:  $Jinqián \sqsubseteq (m:sort \sqcap pl:quant) \sqcup (m:sort \sqcap mass:quant)$ ,  $change \sqsubseteq m:sort \sqcap mass:quant$ ,  $coins \sqsubseteq m:sort \sqcap pl:quant$ . Thus, different NP tokens and respective types can refer to the same type of VOs (and matching token), depending on how the lexicon of a language can establish reference relations to VOs in a context.

Third, the emergence of “thinking for speaking” effects in e.g. lexical decision tasks is expected (Casasanto, 2008). For instance, a bilingual speaker of English and Mandarin may use different strategies to talk about coins in a pocket. While in English two mappings are available (*coins* and *cash*), in Mandarin *Jinqián* can trigger a type resolution task for type (and reference) disambiguation. Different production times are observed, in such cases.

In conclusion, our account offers an empirically sound model on the relation between VOs and NPs. Via the notion of infomorphism, our model can also solve our three problems in a simple and accurate way.

#### References

- B. Abbott. 2010. *Reference*. Oxford: OUP.
- M.C. Baker. 2003. *Lexical Categories: Verbs, Nouns and Adjectives*. Cambridge: CUP.
- J. Barwise and J. Seligman. 1997. *Information flow: the logic of distributed systems*. Stanford, CA: CSLI press.
- I. Biederman. 1987. Recognition-by-Components: A Theory of Human Image Understanding. *Psychological Review* 94(1): 115-147.
- D. Casasanto. 2008. Who's afraid of the Big Bad Whorf? Cross-linguistic differences in temporal language and thought. *Language Learning* 58(1): 63-79.
- T. Gao, G.E. Newman and B.J. Scholl. 2009. The psychophysics of chasing. *Cognitive Psychology* 59(2): 154-179.
- P. Pietroski, T. Lidz, T. Hunter & J. Halberda. 2009. The meaning of ‘Most’: semantics, numerosity and psychology. *Mind and Language* 24(5): 554-585.
- D. Marr. 1982. *Vision*. The MIT Press.
- F. Landman. 2000. *Events and Plurality: the Jerusalem Lectures*. Dordrecht: Kluwer
- T. Metzinger and V. Gallese. 2003. The emergence of a shared action ontology: building blocks for a theory. *Consciousness and Cognition* 12(4): 549-571.
- J.I. Olszewska. 2011. Spatio-Temporal Visual Ontology. *Proceedings of the 1<sup>st</sup> EPSRC/BMVA Workshop on Vision and Language*. Brighton, UK.
- F.J. Pelletier. 2009. A Philosophical Introduction to Mass Nouns. In F.J. Pelletier, editor, *Kinds, Things and Stuff*, pages 123-131. Oxford: OUP.
- S. Peters and D. Westerståhl. 2006. *Quantifiers in Language and Logic*. London: Clarendon Press.
- B.J. Scholl. 2007. Object persistence in philosophy and psychology. *Mind & Language*, 22(5): 563-591.
- B. Smith. 1995. Formal Ontology, Common Sense and Cognitive Science. *International Journal of Human-Computer Studies* 41(8): 641-66.
- M. Yamamoto. 1999. *Animacy and Reference: a Cognitive Approach to Corpus Linguistics*. Amsterdam: John Benjamins.
- S. Zhu, S., Y. Wang and Z. Xu. 2005. What are textons? *Int. Journal of Computer Vision* 62(1): 121-131.