Possible Nouns for Visual Experiences: A Theory of the Vision-Language Interface

Francesco-Alessio Ursini
Sun Yat-Sen University

Paolo Acquaviva
University College Dublin

FRANCESCO-ALESSIO URSINI
Sun Yat-Sen University

PAOLO ACQUAVIVA
University College Dublin

POSSIBLE NOUNS FOR VISUAL EXPERIENCES:
A THEORY OF THE VISION-LANGUAGE INTERFACE

ABSTRACT: The relation between vision and language is ana-
lyzed through a formal statement of what defines objecthood in
the two domains. An interpretation of independently-motivated
approaches to vision and to the grammar of nominals allows us to
define the connection between them as an "infomorphism" con-
sisting of two functions. Visual and linguistic objects are only in-
directly related: the functions range over types and tokens, whose
map defines objecthood in each domain. We show how the infer-
ences proved in this system are empirically correct, and we draw
some conclusions about the import of our proposal on the role of
language in cognition.

1. INTRODUCTION

Language is so fundamental in representing what we think, that a major
challenge for studies of cognition and conceptual knowledge is avoiding
the common pitfall of equating concepts with lexical concepts (Kelter
& Kaup 2012). Usually, the strategy for avoiding this pitfall consists

in devising experimental tasks that bypass language and allow a direct investigation of conceptual content independently of its verbalization (Papafragou 2005). We explore here a different direction, and compare the conceptualization of objects in language and in vision.

Our first aim is to make such a comparison possible in the first place. This target is far from trivial, because "the conceptualization of objects" is too vague a notion for investigation unless it is operationalized precisely enough to allow a comparison of different ways of defining objecthood. For that purpose we propose a formalization of the visual and nominal conceptualization of objects in terms that have theoretical and empirical justification by themselves (Sections 2 and 3), relying on the extensive body of research made in these two intensely studied systems. The central aspect of our proposal is the formulation of a formal system of correspondences between the two, based on the notion of infomorphism (Barwise & Seligman 1997) and implemented by means of two functions (Section 4). As we show in Section 5, this allows us to formally derive, as necessary inferences, several different matches between linguistically defined entities (nominals) and visually defined ones (visual objects). Importantly, the formal inferences cover the *absence* of visual input or of recognized linguistic material, and above all, they predict what correspondences are not definable.

We will show that these results are all empirically supported. As will also become clear, our analysis does not need to impose an otherwise unmotivated isomorphism between nominals and objects. Nominal linguistic structures and objects of visual representations are commensurable abstract entities thanks to our formal interpretation of independently justified properties, not because we present them as parallel to begin with. Still, while being neither isomorphic nor derived from each other, vision and language emerge from our analysis as two representation systems whose interaction can be investigated in an empirically verifiable and predictive way.

2. VISUAL OBJECTS: AN ONTOLOGY OF VISION

Our first goal is to outline an ontology of vision. We proceed in two steps: we first discuss previous research on visual types allowing the recognition of objects (Section 2.1), and then offer a formal treatment

of the corresponding ontology (Section 2.2).

2.1. *Visual Objects: Basic assumptions*

The human visual system processes and segments the visual stream into spatio-temporal units (Michotte 1963, Ullman 1979, Marr 1982, 10–34). These units can represent "objects" for "things" in the world, if they are related to mental categories. Several sub-systems participate in parallel to this process; their governing principles seem constant (Scholl 2009). Thus, even if shape and colour are types of information computed by different sub-systems, a blue ball as an object can be defined by integrating this information.

Models of object recognition abound, although they sometimes offer opposing views as to how these processes occur. Here we take a computational view; we follow the *Multiple Object Tracking* framework of reference (Pylyshyn 1984, 34–40, 1989). In this model, discrete units of sensory experience ("visual stimuli") can be classified according to the visual properties they realize. A visual stimulus is modeled as a *token* or particular instance of a general category or *type*. Stimuli and properties stand in a token/type relation of *instantiation*. The visual index for this relation representing mind-external objects is known as a *Finger of INSTantiation* or FINST (Pylyshyn 1989, 69; Scholl et al. 2001). FINSTs act as "mental fingers" that instantiate objects, in turn represented as relations between a visual token and the types it instantiates (e.g. "blue" and "ball" in our example). Beyond the Multiple Object Tracking framework, FINSTs are known as *object files* (Kahneman et al. 1992; Zacks & Swallow 2007).

FINSTs/object files, thus defined as relations between tokens and types, can become spatially and temporally stable in short-term memory when they meet certain constraints: occupy connected positions (connectedness), consist of elements mutually connected by a part-of relation (mereology), remain stable over time unless external changes occur (persistence), and never result in conflicting representations (consistency) (see Scholl 2001, 2009). Observers can abstract complex categorial representations, which become stable units in long-term memory over repeated exposure.

The object classes so defined are known as object *concepts*, and represent shared information about object files and their relations, stored

in long-term memory (Biederman 1987; Cowan 2005). For instance, the "bicycle" object concept can correspond to a representation for any object having certain key parts (two wheels, a frame) and functions (use as a locomotion tool). Observers can use this tacit knowledge to complete and update these representations. Thus, an observer witnessing a bicycle missing a wheel will still be able to recognize this object as a bicycle (see Scholl & Pylyshyn 1999; Serre et al. 2005; Feldman 2007).

Works on possible visual types abound, and do not necessarily offer a cohesive picture. Still, they provide extensive evidence that the classification process follows a type-driven principle. Examples of types include *textons*, which include material constituency (e.g. a steel container) and colour (e.g. a red ball: Julesz 1981; Zhu et al. 2005). *Shape* types have also been studied in detail (e.g. Marr 1982; Biederman 1987; Hummel & Biederman 1992; Riesenhuber & Poggio 2002). Quantities or *units* are another type, and can be classified according to their quantity (e.g. "nine" objects: Pietroski et al. 2009) or organization. Thus, several dots moving in the same direction can also be individuated as a single complex unit or group (Scholl et al. 2001; Pylyshyn 2004, 2006; Mitroff et al. 2005).

Objects can also be individuated via their potential functions for a subject in an environment, or *affordances* (Gibson 1966; 1979, Ch. 1; Wells 2002; Withagen et al. 2012). *Motion* is also conceived as a visual type that allows agents to identify objects against static backgrounds (Ullman 2000; Gao et al. 2010; Scholl & Gao 2013). Affordances and motion types are sometimes conceived as more "intensional", or "higher order" visual types (cf. Wells 2002; Scholl & Gao 2013). We focus on these visual types as the starting points for such inferences, and leave the discussion of inferences on non-visual information for future research. Here, we proceeded by outlining a formal ontology of visual objects.

2.2. *Visual Objects: A formalization*

Multiple Object Tracking theory offers a formal view of these notions, although certain aspects are not fully developed (e.g. concepts). Here we follow this formalization and expand on the aspects that are underdeveloped, to pave the way for our proposal in Sections 4–5. Formally,

a FINST/object file is defined as *FINST*(*token*,*type*)=*d* (Pylyshyn 1989, 75–76). The index *d* acts as an index (or "mental finger") representing the object as a relation between a visual token and the types it instantiates. The next step is to define how we represent object concepts. This aspect is usually underdeveloped in the theories discussed so far, but can be made precise by implementing a structured type theory that captures how visual types can be combined into complex types. Here we adapt the core ideas of Type Logical Composition (Asher 2011).

A simple way to present how this theory works is as follows. A bicycle can be conceived as an object that involves different parts with different shapes and colours (e.g. a black frame, red wheels, and so on). Thus, an object can be represented as a combination of parts that can be recognized in their own right, but that also combine to form a bicycle as a whole. Several works follow logic-driven, algebraic models of representation to capture these facts with respect to one type of information, as we have discussed (e.g. Marr 1982, 225–230; Hummel & Biederman 1992; Pylyshyn 2004). Type Logical Composition follows a similar tack, and proposes a formal system to capture how these distinct types are combined. A central aspect of this proposal is that this model of types is identified with a *Boolean Algebra*. This algebra is defined as a structure in which the three operations *meet*, *join* and *negation* and the relation *part-of* are defined. To make the discussion maximally transparent, we represent this and the other operations/relations by representing how types can be applied to tokens.

The operation *meet*, represented via the symbol "⊓", is defined as an operation that takes types and returns their common type(s), defined as "$a(x) \sqcap b(x)$". Thus, $blue(x) \sqcap ball(x)$ is the meet of the visual types that represent a blue ball, via a visual token x.[1] The operation *join*, represented via "⊔", takes types and returns their unified type(s), defined as "$a(x) \sqcup b(x)$". Thus, an ice-cream shape corresponds to the join of two more basic shapes, that of a cone and a sphere: we have $cone(x) \sqcup sphere(x) = ice\text{-}cream(x)$. The operation *negation* "¬" is defined as an operation that takes a type as an input and returns its complementary type as an output. If a square is divided into two distinct parts, one black and the other white, then the colour type of one part will be the complementary type of the other part. We thus have $black(x) = \neg white(x)$. The *part-of* relation "⊑" (i.e. "$a(x) \sqsubseteq b(x)$") is de-

fined when the two identities "$a(x) \sqcap b(x)$" and "$a(x) \sqcup b(x)$" hold. That is, type a is *part-of* type b if their *meet* is the (sub-)type a, and their sum is the (super-)type b. A second relation is the identity relation. Two object files are identical (i.e. we have $a(x)=b(x)$) if their meet and join are also identical (i.e. we have $a(x) \sqcap b(x)=a(x) \sqcup b(x)$).

Let us now turn to object files. According to these definitions, an ice-cream object file, call it b, can be defined as $b=(cone(x) \sqcup sphere(x))$. An index b is identified with an instantiation relation between a visual stimulus x and a join of two shape types, *cone* and *sphere*. A similar analysis can be extended to meet types. For instance, a blue ball can be defined as the object file $d=(blue(y) \sqcap sphere(y))$. The object file d is identified with a visual token y, instantiating the visual types *blue* and *ball* at the same time, hence the use of the meet operation. For the role of negation and the part-of relation in object files, we defer a discussion when we will tackle the relevant examples.

Consider two object files for "balls", $d=(blue(y) \sqcap sphere(y))$ and $f = (red(z) \sqcap sphere(z))$. These files differ with respect to the colour they instantiate, but both can stand as specific files related to the concept of "ball". In this case, we say that the colour type of the concept is *underspecified* with respect to other types; one value assigned to a type can be ambiguous in context (cf. Pylyshyn 2004; cf. Asher 2011, Ch. 3 for a linguistic definition). The shape type can also be underspecified, since balls can come in different shapes, in football codes (cf. the oval ball used in rugby). If d and f are the only two files acting as members of the concept *BALL*, then this concept represents the class of spherical-like objects of any colour.

A concept C can be extensionally conceived as the set of object files instantiating certain properties. Thus, the object concept of red things is the collection of object files o, o', o'',... individuated via the type *red*. We represent object files via lower-case letters distinguished by superscripts. We represent concepts based on types by means of upper-case letters. Thus, the concept "red" is represented as *RED*; $\{o, o', o'',...\} = RED$ is the relation stating that a concept is defined as the set of object files of *red* type. In intensional terms, we represent concepts as relational entities. We have $o \in RED$, with "$\in$" for the *membership* relation, to represent how concepts are formed with object files as building blocks.

The concept *BALL* can be represented as the join *BALL*=(*SPHERE* ⊔ *COLOUR*), with *SPHERE*, the concept, derived from type of spheres, and *COLOUR* being the complex (join) concept of all *colour* types (e.g. we have *col*=*red*⊔*green*⊔*blue*..., hence *COLOUR*=...). For the sake of consistency, we employ the variant notion used in the literature on Boolean operators to represent concepts (e.g. Landman 1991, Ch. 2–4; Chierchia 1998 on kinds). We thus have *CONCEPT* = ⌊⌋*file*: an object concept is the join of all files representing a given object type (e.g. *ball*(*x*)⊔*ball*(*y*)⊔*ball*(*z*)).

Two key results emerge from this initial formalization. First, concepts are defined insofar as abstraction from more than one file is possible. *Beagle* and *dachshund* are "types" of dogs defined via their common characteristics (e.g. being four-legged) but also their differences (e.g. size and coat). Our system proves that the concept *DOG* represents the combination (union) of more specific objects or "variants" that qualify as dogs because of their similar, and yet distinct visual types. Second, FINSTs/object files internally represent mind-internal information about mind-external "objects" in a bottom-up manner. Concepts can develop and change over time via information obtained via new object files (cf. Laurence & Margolis 1999; Margolis & Laurence 2007; Murphy 2002, Ch. 2).

Overall, we have shown that the notions of objecthood, visual object files and concepts support a conception of vision as a mediated process, which starts from the classification of tokens (visual stimuli). The mental ontology of visual objects reflects mind-internal principles regarding how this information is processed and organized. Object concepts act as mental objects that guide this classification process even if the visual input involves apparently incomplete or underspecified information.

3.   NOMINAL OBJECTS: AN ONTOLOGY OF LANGUAGE

Moving from the visual to the linguistic conceptualization of objects, we meet an initial problem; semantically, objects are simply presupposed, as a domain over which the denotation of linguistic expressions is defined. For example, *ball* can be (and generally is) ultimately analyzed as an expression true of balls, *red* of red things, and *see* of pairs of objects where one sees the other. A comparison with the visual system

requires a different perspective, namely asking how language *defines* and *characterizes* objects. While both pronouns and nouns categorize the entities they denote (by means of gender and number features, for example, or in terms of pragmatic role for some honorific pronominal forms), it is specifically nouns that describe them and thus define an ontology by linguistic means. Accordingly, we discuss the special role nominals have in in representing types of entities (Section 3.1), and indicate how linguistic determinations contribute to this representation (Section 3.2). Finally, Section 3.3 interprets the definition of objects in language in type-token terms that allow us to relate it to their definition in vision.

3.1. *Nouns and nominals as expressions for abstract types of entities*

Nouns are one of the fundamental lexical categories, and so the question of how they can define objects intersects the rich literature on the content, definability, and universality of lexical categories (Croft 1991; Baker 2003; Gil 2005; Bisang 2011; Panagiotidis 2014; among many others). However we are not (directly) interested here in the possible semantic latitude of nouns as a category of words, but more precisely of nominals, understanding under this term single lexemes like *wine*, complex structures like *black hole* or *footwear*, or phrases like *good dancer* or *a bottle of ordinary wine*. Nominals in all of these 'sizes' define object types.

This circumstance accords naturally with a theoretical approach which views nouns (and all lexical items) as constructions organized on the same principles that underlie syntactic constructions (see Borer 2005a, 2005b; Harley 2014; cf. Rijkhoff 2002 for a different approach). Our proposal presupposes this theoretical perspective, and so sits most comfortably within a non-lexicalist approach to morphology and syntax; however, it does not strictly require this theoretical choice (which will only be crucial for the interpretation of uninflected lexical bases as existing linguistic objects, cf. 5.2 below).

On the other hand, a semantic assumption is central: that nouns and certain nominals like *red wine* denote not just the mind-external objects that they are true of, but also, and indeed primarily, an abstract kind which they identify and name. We thus follow Krifka (1995) and Zamparelli (2000) in assuming that nouns primarily denote an abstract

kind, and are then specified by morphosyntactic determinations (like number or quantificational expressions) which turn them into predicates true of entities (see Acquaviva 2018 for details). It is in this sense that nouns and nominals *define* object types, and are not just true of already given elements of the domain. A noun like *ball* and a kind-identifying phrase like *red wine* denote, then, a kind-level, abstract object, which is part of their semantic interpretation and acts as an intermediary for the reference to one (for *ball* singular) among a set of mind-external balls, or to a concrete portion/instance of red wine, or possibly to the abstract kind 'red wine'. This abstract object is part of the hypothesized semantic interpretation, and in this sense it is a language-internal construct; whatever its possible relation with a concept, it is not here posited as a psychologically defined concept, in particular not as the concept of an entity that exists in the ontology presupposed by the speaker.

To see what this means, consider nouns like *child* or *president*: they define roles, by definition not identical with individuals (if John is a judge and a janitor he is not two people), and so the corresponding kinds cannot identify individuals as members of the domain. However, these roles too have individuality, in the sense that the approach we follow reifies them as abstract sorts, on a par with a sort like *the average American*. We follow Collins (2017) and accept that this description denotes an abstract object which can be used as an argument of a proposition that is true or false in the world (like *The average American has 2.3 children*), even though by itself the object is abstract, does not denote any concrete mind-external individual, and speakers are aware of that.

3.2. *Defining and characterizing kinds of objects*

Nouns and nominals define and characterize objects insofar as they describe them in different possible ways. *Home* and *house* have largely (not totally) overlapping extensions, but they characterize in different ways what they are true of; a *holiday house* can be termed *holiday home* to present it in a more attractive light, but the protective warmth associated with *home* makes it unsuitable for use in, for instance, *jailhome*, *workhome* or *whorehome*. These may look like 'lexical' differences that do not affect the 'grammatical' identity of the two words (here, nouns)

as linguistic objects. But the facts stand otherwise. *Shadow* and *shade* or *sun* and *sunshine* exemplify pairs (not minimal) where the second member strongly favours a reading where the denotation has the structure of an unbounded continuum, not a set of individuated wholes. The point is that this distinction is made visible by a difference in syntactic distribution; we are of course alluding to the contrast between so-called mass and count, which is encoded in grammar and emerges as a distributional asymmetry. The same happens, more sharply, with well-known pairs like *hair/ hairs* and *hair*, or with the different readings of *cloud/ clouds* (*a cloud*, *a few clouds*) vs. singular mass *cloud* 'cloud cover' as in European English *there will be sunshine through broken cloud*.

Differences in countability properties notoriously emerge across languages, highlighting different conceptualizations for nouns that would otherwise seem to denote the same referents. The English *grape/ grapes* describes a certain kind of fruit as a collection of berries; Italian describes the same fruit as *uva*, a singular noun with the mass syntax of a substance term like *acqua* 'water'. Similarly, what English describes by the plural count noun *peas* is described in Russian by the singular *goroh*, which has the syntax of a mass noun like many other denominations for what is seen as a granular substance (cf. also the Russian singular *luk* 'onions'). To describe a single pea, Russian has a derived word, *gorošina*; this is however a distinct lexical item, with its own plural *gorošiny*. So-called singulative derived nouns play a much more central role in the grammar of languages like Arabic, where *zahr* 'flowers' contrasts with the singulative *zahra* '(a) flower', and the latter can be pluralized as *zahraat*; in addition, the 'plural' base itself can be pluralized in turn, as *zuhuur* or *azhur*.

In each of these forms a grammatical characterization identifies a particular type of interpretation. As in a myriad other cases, grammar affects the way a noun characterizes its denotation. Cross-linguistic studies have extensively documented how semantic interpretation varies as a function of different choices of number, gender, noun class, or classifier (Massam 2012; Senft 2000; Aikhenvald 2003) and we will not attempt to summarize them here. What this evidence shows is that nouns and complex nominals are not just true of objects whose nature is given; languages characterizes the nature of these objects by distinct lexical choices (*home* vs. *house*, *wood* vs. *timber*), by distinct complex

structures (*red wine* vs. *mulled wine*), by distinct grammatical choices (Swahili *mi-ti* 'wood', class 3 vs. *ki-ti* 'chair' class 7; Italian *legno* masc. 'wood' vs. *legna* fem. 'firewood'). We therefore include all of these language-internal representations as means to linguistically define and characterize objects.

3.3. Types and tokens

The way language defines objects is thus very different from the way vision interprets and categorizes visual stimuli. Our goal of examining the relation between the two requires that they are made comparable, but not by forcing an isomorphic structure on them. Our characterizations of the definition of objects in vision and in language are both independently supported by a rich literature, and obliterating them to impose an isomorphic structure on both would strip the argument of its empirical justification.

We therefore take a different approach. This section has already spelled out our assumptions about the grammar and semantics of nouns. Beyond that, there is no need to formulate and adhere to a particular grammatical interpretation of the structure of nominals. However a noun or a noun phrase is grammatically analyzed, it is possible to distinguish the structure in itself from the choice of morphemes or lexical items (this depending on one's assumptions about morphology) that flesh it out. For example, assigning for simplicity a basic analysis as in (1a) to the nominal the lazy dog, it is possible to abstract from it the structure (1b), and to list as in (1c) the lexical/morphological material that interprets it:

(1)   a. $[_{Det}$ the $[_{Adj}$ lazy $[_N$ dog $]]]$

　　　 b. $[_{Det}$ Det [DEFINITE, SG] $[_{Adj}$ Adj $[_N$ Noun [SG] $]]]$

　　　 c. *the, lazy, dog*

Many questions remain open from a linguistic perspective: for instance, whether the English grammatical morpheme *the* is entirely determined by the features 'definite' and 'singular' (in fact, whether these features are justified, since 'singular' could conceivably be [−plural], or the default value in a privative feature). In that case, *the* would not be selected among competing lexical choices in the same sense in which

lazy and *dog* are, because the featural content would entirely determine its choice (in other words, the choice of 'grammatical' morphemes follows deterministically from the choice of features, unlike the choice of 'lexical' morphemes or lexemes). Likewise, the labels 'adjective' and 'noun' for *lazy* and *dog* could conceal a more complex structure where lexical roots are grammatically categorized by a categorizing null formative or as a function of their syntactic placement (see Borer 2005a, 2005b; and Harley 2014). Finally, the nature of *dog* as a singular count noun is likely to be reflected in a finer-grained structure, which includes a separate locus for number features and/ or for the determination of countability. These and similar questions, no matter how substantive in themselves, are irrelevant here. What counts is that we can distinguish two types of linguistic representations of a grammatical object, both of them abstract. (1b) is abstract because it simply represents a template for a type of linguistic object, not yet realized by any morpheme choice to interpret it. But (1a) is also abstract, because the morpheme choices indicated make up a *type* representation, not a *token* utterance. (1a) realizes (1b), but it is still an abstract representation because the morphemes indicated as *the*, *lazy*, and *dog* are abstract symbols standing for any suitable realization (in that structural context). This helps clarify the status of them when they are listed in (1c), abstracting away from the structure they flesh out in (1a); here, they represent abstract symbols for elements that only appear in a grammatical representation (like (1a)), but are here listed as an unstructured array. They are not exponents, phonological matrices, or semantic building blocks; they are just symbols chosen from the lexicon of one particular natural language. Speakers of English have one definite article notated *the*, unlike the series of grammatically differentiated inflected articles of German, for example, and abstracted from the structure in which they appear. Their identity with orthographic words is of course purely a matter of convenience.

To sum up: we factor out an uncontroversial representation like (1a) into a pure structural representation (1b) and a choice of vocabulary items (specific to one language) which interpret it lexically. Both representations are abstract, but their mutual relation is one of a specific choice of linguistic material (morphemes, lexemes) realizing a grammatical template; this can be construed as a type-token relation,

as we will now make precise.

4.   THE NOUNS-OBJECTS INTERFACE: A HYPOTHESIS

We start from the presupposition that language and object recognition are distinct modules of cognition, and propose a formal apparatus to capture their relations and reciprocal constraints (Section 4.2). This apparatus is based on Barwise & Seligman (1997)'s theory of information flow. Before we spell it out, we motivate our choice (Section 4.1).

4.1.   *Motivation*

Given our presupposition about language and object recognition, our account of the relation between these two aspects of cognition is incompatible with a certain class of theoretical frameworks. Works within the "Construction Grammar" programme assume that there is no distinction between the cognitive principles regulating language and other domains (e.g. Goldberg 1995, 2006; Talmy 2000; Tomasello 2003; Evans 2010). They would thus predict that no relations would arise between these domains, for no such distinction would be in place. Other works do propose a distinction amongst domains (or "modules", cf. Fodor 1975, 1983), which are usually modeled as involving hierarchical structures (Hauser et al. 2002; Jackendoff & Pinker 2005; Kirby 2007). However, these works generally do not offer accounts of how the relations between language and non-linguistic domains operate. Thus, different linguistic theories converge in remaining vague about these "interface" matters (cf. Chomsky 1995, 2001; Sadock 2012; Sag et al. 2012; Harley 2014). A partial exception is Jackendoff (2002), in which modules processing information in parallel can involve forms of "binding" (i.e. synchronization) of their processed information.

   A red thread connecting these theories is that, even when mappings between domains are presented, their properties are not explored and formalized in detail. Our goal in using the theory of information flow (Barwise & Seligman 1997) is to remedy this situation, thereby showing that the "logics" of visual objects and syntactic structures affect one another in non-trivial ways.[2] We have three reasons for our choice. First, the formal apparatus that the theory offers allows us to capture cognitive and linguistic aspects with ease, while previous formally oriented

works (e.g. Gupta 1980; Lasersohn 2018) mostly focus on language alone. Second, the formal component spells out explicit definitions of notions such "type", "token", "relation", "classification", and so on (Barwise & Seligman 1997, Ch. 2–4; cf. also Smith 1995, 2003). Third, the system offers two precise inference systems that allow us to define what counts as a type in both domains, and to make formally precise the notions of object file, concept, nominal and kind discussed in Sections 2–3. With these clarifications in mind, we turn to the proposal.

4.2. *The Apparatus*

An information system can be envisioned or displayed via the logical properties and structures that govern the types constituting this system, and their corresponding relations. A "logic" of visual objects must thus capture how different visual types combine to identify objects files, in turn constituting object concepts. A "logic" of nominals must capture how different nominal grammatical types classify specific lexical choices. The two systems or "logics" can be defined via the operations that govern these properties, and via an *infomorphism*, a mapping between the two systems that govern their information exchange. Intuitively, we can conceive each formal system as a tool that allows us to study object recognition and lexical interpretation as "parts" of visual cognition and language. These two larger parts can then be conceived as parts of cognition as a whole.

The infomorphism captures how the two systems must exchange information for this exchange to be successful: how nominals can describe and define objects, and how visually defined objects can be labelled via the linguistic representations known as nominals.

We define the formal properties of each system via the compact formulations in (2)–(4):

(2) A *Visual Structure* (*VS*) is a sextuple $<O, *O, \models, \sqcup, \sqcap, \sqsubseteq>$ consisting of:

- A set of tokens O
- A set of types $*O$
- An instantiation relation $\models$
- The two operations $\sqcup$ (join), $\sqcap$ (meet)

  – The relation ⊑ (part-of)

(3)   A *Nominal Structure* (*NS*) is a sextuple <*N*, *N, ⊨, ⊔, ⊓, ⊑> consisting of:

  – A set of tokens *N*

  – A set of types *N

  – An instantiation relation ⊨

  – The two operations ⊔ (join), ⊓ (meet)

  – The relation ⊑ (part-of)

(4)   Variables:

  – *o* is a member of *O*   (a visual token)

  – *o is a member of *O   (a visual type)

  – *n* is a member of *N*   (a nominal token)

  – *n is a member of *N   (a nominal type)

The two structures act as *classifications* or *classifying structures*: they allow us to classify visual and nominal entities as combinations of tokens (from the sets *O*, *N*), types (from the sets *O, *N), and relations among these types. Visual tokens can be conceived as visual stimuli that can be interpreted as realizing visual types (e.g. *red*, *ball* and so on). Nominal tokens are linguistic symbols without their grammatical content. The *instantiation* relation "⊨" represents a relation between tokens and types. We say that *a* instantiates *b* (i.e. $a \models b$) when a (nominal, object) token *a* acts as an instance of the (nominal, object) type *b*. No confusion should arise regarding this symbol. Although its use in propositional logic is to model entailment relations among propositions, in our case it establishes an "asymmetrical" relation between an "abstract" type/ property and its "concrete" realization/ token (Barwise & Seligman 1997, Ch. 3). Since the notation *type*(*token*) in Sections 2–3 is equivalent to this relational notation, we use it here onwards for consistency and readability reasons.

   In our formal system, object files (e.g. *b*=*ball*(*x*)) and nominal representations (e.g. [*the* [*dog*]]) are not basic but involve the definition of complex types via the two operations ⊔ (join), ⊓ (meet) and the relation ⊑ (part-of), introduced in Section 2. We thus assume that

an object file can be defined as instantiating either the meet of two or more types (e.g. $red(x) \sqcap sphere(x) = ball(x)$), or the join of types (e.g. $red(x) \sqcup blue(x) = colour(x)$, in our simplified model). Portmanteau affixation in the nominal domain involves the meet of nominal types: the Italian root noun *can–* can be suffixed via the inflectional suffix *–e*, which fuses the *singular* and *male* (*gender*) types (i.e. we have $singular(x) \sqcap female(x) = can–e$). Join nominal types involve a form of "ambiguity" (polysemy, as we discuss in Section 5). Thus, English *fish* is ambiguous with respect to number (i.e. it can be $singular(x) \sqcup plural(x)$).

The $\sqsubseteq$ (part-of) relation allows us to capture type relations. For instance, $blue(x) \sqsubseteq colour(x)$ and $singular(y) \sqsubseteq number(y)$ model relations between object and nominal types, respectively. To an extent, this relation captures the fact that a certain type can be "part" of a larger domain of types, as the examples suggest. At the same time, this relation indirectly suggests that object files are related to concepts as "parts" of a complex representational structure based on a type (i.e. *COLOUR* including *RED* and *BLUE* as sub-concepts). The same reasoning applies to nominal categories and their possible declension classes (e.g. *ball*, *balls* as specific forms of the nominal class in English).

Overall, our classification structures offer a logic that makes possible a formal proof of what count as visual objects and nominals in the first place. The second inference system is the infomorphism. To see how this system works, we present a graphic representation in Fig. 1, and explain how this representation should be read:
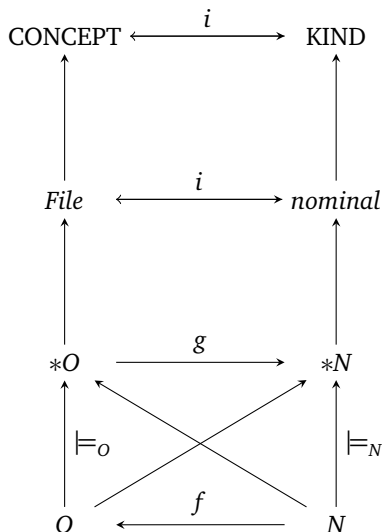
Figure 1. The structure of our infomorphism.

The graphic reads as follows. The instantiation relation $\models_O$ represents a relation holding between object tokens and types; the relation $\models_N$ that between nominal tokens and types. The first instantiation relation defines object files; the second, grammatically inflected nominals. Both relations can be conceived as relations activated in short-term memory.

We can now offer the central definition. Given our classifications *VS* and *NS*, an infomorphism is defined as a pair of functions $i=<f, g>$ such that:

(5)   a. $f(n)\models$ *o if and only if $n\models g($*o$)$

     b. *f* establishes a mapping from *N* to *O*

     c. *g* establishes a mapping from *O to *N

In words, *f* and *g* are symmetrical functions (mapping relations) connecting objects and nominals. The function *f* maps nominal tokens to visual tokens; the *g* function maps visual types to nominal types. Thus, the general equations $f(n)=o$ and $g($*O$)=$*N hold (cf. Barwise & Seligman 1997, Ch. 3). These general equations are represented as

the diagonal lines connecting tokens from one system to types of the other system. At the level of files and nominals, an infomorphism *i* establishes an indirect relation inferred from *g* and *f*. The infomorphism represents the fact that these two entities are related, but the precise relations between constituting tokens and types become opaque. A clear consequence, fully discussed in Section 5, is that the correspondence between nominals and files, and therefore concepts and kinds, does not involve nouns or nominal structures of a specific "size". The equations *i(file)* = *nominal* and *i(CONCEPT)* = *kind* thus represent these indirect, "horizontal" mappings, in which concepts indirectly constrain kinds and vice versa. The same reasoning applies to files and nominals, given this mapping.

5. THE INTERFACE: A THEORY OF INFORMATION FLOW

This section articulates the consequences of the formalization just proposed. The two functions posited between the visual and the linguistic domain constitute a map that is systematic but indirect, which is empirically correct (Section 5.1). The most instructive cases, and those which most clearly demonstrate the empirical success of our proposal, are those involving non-defined and zero values (Sections 5.2 and 5.3 respectively).

5.1. *The virtues of an indirect mapping*

Our hypothesis is centered on the assumption that the visual and nominal domains are connected via a homomorphism, a "many-to-one" mapping (Landman 1991, Ch. 2; Barwise & Seligman 1997, Ch. 2). This simple formal aspect of our hypothesis already entails two important empirical consequences.

Recall that, first, we have the mapping $f(n) = o$, with n a variable over lexical material (words or morphemes), and o a variable over visual stimuli; and second, the mapping $g(*o) = *n$, with $*o$ a variable over visual types, and $*n$ a variable over linguistic types. Since the function f is many-to-one, it can map several alternative lexical choices over a visual stimulus (i.e. name an object in different ways, e.g. *ball* vs. *sphere*), which is obviously the desired result: a visual experience does not dictate a unique choice of words that describe it. Likewise,

the function g can map different visual features (e.g. being spherical, being blue) to a single grammatical structure, which again is the desired result. One grammatical nominal structure can underlie the description of many different visual features. That visual and linguistic objects clearly do not deterministically select each other is so obvious that it obscures the fact that our system derives it while still establishing a precise relation between the visual and the linguistic system. The infomorphism depicted in Fig. 1 achieves this result because the two functions f and g directly connect only visual tokens to grammatical structures and, respectively, lexical tokens to visual types (features). In both domains, an object is defined as a type-token map, but there is no direct connection between objects as thus defined. The only "global" connection between them is represented by the composite function i, which is defined as the pair of functions $<f,g>$. This indirect formal link ensures that concepts abstracted from visual object files and kinds associated to linguistic nominal representations are related, avoiding the pitfall of making visual and linguistic representations directly dependent on one another. At the same time, the infomorphism is flexible but far from vacuous: a visual input (token) categorized by the features (type) *BLUE⊓BALL* can never be related to a choice of lexical material (token) *red* and *ball*.

To see this, let n stand for the lexical array {*red*, *ball*}, and the meet type *BLUE⊓BALL* for the visual type $*o$. By hypothesis, n does not instantiate $g(*o)$; that is, {*red*, *ball*} does not instantiate the grammatical template [*Adjective* [*Noun*]] which is related by function g to the type *BLUE⊓BALL* ($g(*o)$). Since it is not the case that $n{\models}g(*o)$, (5a) says that $f(n){\models}*o$ will not hold either, that is, the function f will not map {*red*, *ball*} to a visual stimulus that instantiates the $*o$ type *BLUE⊓BALL*. In other words, no matter how we characterize the visual type notated *BLUE⊓BALL*, if it is not what language expresses by the words {*red*, *ball*}, then those words cannot be part of a linguistic description of the visual stimulus categorized as *BLUE⊓BALL*. The fact that visual types must be labelled verbally, like *BLUE*, might raise the suspicion that we are trivially stating that what is "red" cannot be "blue". Not so: the nature of visual types is established vision-internally. Still, our system is not vacuous because it rules out a connection between a nominal and a visual object, unless the functions f and g ensure that a token of one

domain can instantiate a type in the other.

The infomorphism defined between the two domains, however, is not just empirically adequate, but also revealing, since it derives correct results that would not be immediately obvious. To see this, we first discuss cases involving undefined and null values for variables and types (Sections 5.2–5.3). We then discuss how "regular" values can influence the emergence of well-known phenomena, such as prototype effects, polysemy and relativism (Section 6).

5.2. *Undefined values*

Let us consider cases in which tokens and types take an undefined value, represented as "⊥". At a pre-theoretical level, this is true in two cases: a token/ type not being previously defined. The first is when tokens do not act as the minimal units in either domain (e.g. visual stimuli not being spatio-temporally discrete, lexical entries being novel). The second is when types are not identified that can determine how the instantiation process occurs (e.g. visual stimuli having a novel, not previously defined shape, nominals lacking inflectional morphology).

At a token level, we can have undefined tokens o (objects) or n (nominals). In the first case, the relation $(\perp \models^* o) = \perp$ holds: a visual token is undefined, so it cannot instantiate a visual type. However, the function f can still map a linguistic token onto this undefined visual type (that is, $f(n) \models^* o$ holds). Conversely, the function $n \models g(^* o)$ is defined, and connects nominal tokens and visual types.

This set of relations proves that undefined visual stimuli can still be named via fully formed lexical choices, and via these, it can be established that they are related to visual types. In other words, we predict that we can have nouns for novel objects; that is, nouns whose content is still not defined (e.g. *dax*, Landau et al. 1988, 1992). We also predict that nouns failing to define any content can become part of a nominal system. Forms like *contents*, *beginnings*, *furnishings* are true of objects in a particular (often contingent) spatiotemporal configuration or used for a particular function. However, they do not describe what these objects are in terms of kind instances, even though the plural entails that they make up a collection: with *contents*, for example, we only know that there are entities that are contingently contained, without specifying the kinds they instantiate (Wierzbicka 1988; Acquaviva 2018).

When a nominal token is undefined instead (that is, we have ($\bot \models {}^*n$)), a nominal describing an object file is also missing (i.e. we have ($f(\bot) \models {}^*o) = \bot$). But a token-type relation can be defined in the visual system (i.e. we have $o \models {}^*o$), and with it a mapping between types ($g({}^*o) = {}^*n$). We thus have a proof that not only are non-verbalized visual objects possible, but also that speakers know what grammatical template would be the appropriate one to describe them, if there were the words to describe them. This is what makes novel words possible, for instance when we know that an as-yet undefined property ("having *that* particular look", say) would be an adjective if it was part of the shared vocabulary.

A similar pattern occurs with types. If a visual token that does instantiate a well-defined type (i.e. we have $o \models \bot) = \bot$), it cannot have a nominal to describe it (i.e. we have ($f(n) \models \bot) = \bot$). The same holds for the function g (i.e. we have $n \models g(\bot)) = \bot$). The significant conclusion, formally entailed, is that a visual stimulus cannot be named directly, bypassing its categorization in terms of type, since it does not license the definition of an object file.

Correspondingly, a nominal token not instantiating a nominal type (i.e. ($n \models \bot) = \bot$) cannot serve as a basis for constructing a nominal phrase that describes a visual object. If the token-type instantiation on the linguistic side is undefined, the function g cannot relate a visual type ${}^*o$ to a missing linguistic type (i.e. ($n \models g({}^*o)) = \bot$). This entails that no type-less nominal can name objects, since it is not well-formed (e.g. a root without inflectional morphology, as would be a hypothetical uninflected *gatt–* in a language like Italian). Only fully formed nouns can act as lexical choices which can be related to object files. For the linguistic as well as for the visual domain, then, non-categorized tokens exist, but categorization (that is, an instantiation relation to a type) becomes indispensable for any kind of relation between the two domains.

5.3. *Null values*

A different picture emerges when tokens and types involve a null value, which we denote as "$\varnothing$". We distinguish this value from the undefined value because the null value represents the lack of specific information, rather than the impossibility of defining such a value. For instance, an agent can represent a transparent object as an object lacking a *colour*

type (i.e. we have *colour*: ∅), or a hole as an object having an empty value for the *shape* type. An agent who has not experienced a *blue* colour, however, will represent this first instance as a yet-to-be-defined colour (i.e. "⊥"). The relevant cases are defined as follows.

If a visual token has an empty value, it can nevertheless instantiate a type (i.e. $\varnothing \models^* o$). The corresponding object file can be described via a nominal (i.e. we have $f(n) \models^* o$, $n \models g(^* o)$). If a visual type corresponds to the empty value (i.e. $o \models^* \varnothing$), a similar reasoning holds, but with two slightly different consequences. One can have "null" object files, and therefore concepts, defined via the meet of opposing types (e.g. $sphere(x) \sqcap cube(x)$ as the type of a "spherical cube"). Thus, no object concept can be defined and named as an "impossible" object (see Acquaviva & Panagiotidis 2012). One can also have objects instantiating null types for a given sub-domain, holes in doughnuts being examples (cf. Casati & Varzi 1994). In both cases, the system predicts that the lack of specific information about visual objects, at a token or type level, entails different solutions emerging on how these empty objects can be named.

When nominal tokens and types involve null values, two slightly different patterns emerge. When nominal tokens are "empty" (that is $\varnothing \models^* n$), a nominal lacks a specific meaning in a language, even though it can name an object (i.e. we have $f(\varnothing) \models^* o$, $\varnothing \models g(^* o)$). Thus, English may not include the noun *paccheri* as a noun for a certain pasta shape. However, this noun can be introduced and used in this language to signal this information, even when a speaker explicitly signals that it is not part of his or her current lexicon (e.g. *this is known as "pacchero" in Italian*). If a nominal type is null (i.e. $n \models^* \varnothing$ holds), it can nevertheless describe an object (that is, $f(n) \models^* o$, $n \models g(^* o)$). One can thus have nouns (or pronouns) that involve a neutral or null distinction along a certain grammatical dimension, for instance the neuter of Spanish *esto/eso/aquello* 'this/that', which is neither masculine nor feminine, and does not enter in the agreement relations of the two gender values (Ledgeway 2012, p. 106). Our system thus leads us to posit (again, correctly) that what counts as null may also be a grammatical category value, not just a lexical choice.

6.   THE INTERFACE: EMERGING COMPLEX PATTERNS

We now discuss three phenomena that involve complex relations between linguistic and visual objects: prototype effects (Section 6.1), noun polysemy (Section 6.2), and phenomena related to the "Sapir-Whorf" hypothesis (Section 6.3).

6.1.   Prototype theory

Our infomorphism hypothesis offers a way to model the notion of *prototype* in a relatively streamlined manner. Prototype theory proposes that a prototype is a central concept in a corresponding concept class (Rosch 1973, 1975, 1978). A classic example is that an agent can have the concept *ROBIN* to act as the prototypical concept for the *BIRD* concept class. This is the case when an agent identifies the properties of a robin with the prototypical properties of bird, and possibly associates other *BIRD* concepts with less prototypical statuses. This occurs when a given concept only has some of the properties identifying a prototypical concept: the less ample the set of attested properties is, the less prototypical a concept will be. Prototype theory thus takes a flexible approach to category/ class membership, compared to traditional analyses of concepts (cf. Gärdenfors 2000, Ch. 2, 2004; Murphy 2002).

Our account resembles the proposals offered in *Formal Concept Analysis* (Ganter & Wille 2005; Ganter et al. 2005). We treat *BIRD* as a concept involving (representations of) objects that can instantiate the complex types "'lays eggs, can fly, has feathers'" (cf. Van Eijck & Zwarts 2004, 2). We use partial descriptions for visual properties: "lays eggs" and "can fly" can be conceived as affordances. They capture properties that agents can observe in birds in certain environments, and permit agents to interact with birds (e.g. collect eggs as food). From these properties/ visual types, an agent can develop the *BIRD* concept over the course of multiple observational events. For each event, an object file formalized as $bird = lays\text{-}eggs(x) \sqcap can\text{-}fly(x) \sqcap has\text{-}feathers(x)$ can be instantiated; from multiple files, the concept *BIRD* is formed. An agent usually observing robins as the most common bird species will likely develop the identity $BIRD = ROBIN$ in long-term memory. Under this view, a prototype concept is a concept acting proxy for a super-concept identifying a class, in this case robins for birds, which is the most com-

mon in an environment.[3]

The infomorphism can subsequently establish a relation between nouns for kinds, prototypes and nouns for non-prototypical concepts. In our formulation, the nominal *robin* names this prototype (i.e. we have $g(robin) = lays\text{-}eggs(x) \sqcap can\text{-}fly(x) \sqcap has\text{-}feathers(x)$). Conversely, less prototypical files can be represented as only including part of these types. An agent may observe a penguin as a bird that lacks feathers, and form the object file *penguin* $= can\text{-}fly(y) \sqcap lays\text{-}eggs(y)$ from which the *PENGUIN* concept is formed. *Penguin* becomes the nominal that names this "partial" *BIRD* concept (i.e. we have $g(penguin) = can\text{-}fly(y) \sqcap lays\text{-}eggs(y)$). Our hypothesis can then capture that *PENGUIN* is a less prototypical *BIRD* concept than *ROBIN* via the part-of relation. The concept *PENGUIN* is part of the concept *BIRD*, because the relation $can\text{-}fly(y) \sqcap lays\text{-}eggs(y) \sqsubseteq can\text{-}fly(t) \sqcap lays\text{-}eggs(t) \sqcap has\text{-}feathers(t)$ holds. Since *BIRD* is formed via the concept *ROBIN* and defines the "whole" set of types identifying birds as visual objects, any other bird concept that is formed via only a part of these types will be less prototypical (i.e. we have $PENGUIN \sqsubseteq ROBIN$, which reads: a *PENGUIN* concept is less prototypical than a *ROBIN* concept.

A prediction that emerges from our hypothesis is that agents/ speakers can understand hyponym relations between nouns (e.g. "a robin is a type/kind of bird") as relations between the concepts they are related to (Murphy 2010, Ch. 4; Riemer 2010, Ch. 5). A second, subtler prediction is that no precise relation is established between the types from which a concept arises and the nominal types. Therefore, speakers and speakers' communities may use different nominals to capture this relation. Australians may develop the prototypical identity *PARROT* = *BIRD*, and reason about bird prototypes by using parrots as prototypes, since parrots are more common in this country. A third prediction is that the kinds that nouns denote can also display prototype properties, by being mapped onto concepts. An agent who has *ROBIN* as a prototype concept for *BIRD* will offer *robin* as a name for a prototypical bird, since it denotes the kind associated to this concept (i.e. we have $i(robin) = ROBIN = BIRD$). Thus, our hypothesis reconstructs the central aspects of prototype theory in a compact manner.

6.2. *Noun polysemy*

Polysemy holds when a vocabulary item has when it covers distinct but
related senses (Kearns 2006; Murphy 2010, Ch. 4). For instance, *ball*
has subtly distinct senses, depending on whether one talks about rugby
balls, tennis balls, or other types of objects that may not even be used in
sports. Some cognitively oriented theories propose that concepts and
word meanings/ senses are clearly distinct, although polysemy reflects
conceptual structures (e.g. Tyler & Evans 2003; Evans 2009, 2010,
2015). Other theories do not distinguish the two levels; hence, they
model polysemy as a property of concepts (Murphy 2002; Carey 2009).
Since we model concepts and meanings/ kinds as distinct mental enti-
ties, here we follow the first approach to polysemy.

Model-Theoretic frameworks studying polysemy (e.g. Pustejovsky
1995, 2013; Asher 2011) suggest that polysemous vocabulary items can
carry multiple sense *types*. These determine the possible senses of an
item and its semantic relations with other items. The polysemy of *ball*
can modeled as involving the sense of this item (i.e. $ball(x)$), being part
of at least the types *gender* and *number*. While *ball* is a noun carrying
a *singular* sub-type (i.e. the hyponymy relation $singular(x) \sqsubseteq number(x)$
holds), it is ambiguous (i.e. polysemous) with respect to the *gender*
type. *Ball* can also be considered polysemous with respect to informa-
tion regarding shape; it can be restricted via the adjectives *oval*, *spher-
ical* and so on.

We can represent this fact by representing distinct sense sub-types
as hyponyms of a sense super-type (e.g. $oval(x) \sqsubseteq ball(t)$ and *spheri-
cal*$(y) \sqsubseteq ball(t)$). A type (e.g. *oval*) can be defined as a hyponym/ sub-
kind/ sub-type of another type (e.g. *ball*) if a token belonging to the
first type also belongs to the second type (i.e. $oval(x) \sqsubseteq ball(t)$). The set
of possible hyponyms of a noun can be represented as a join set (i.e. we
have $oval(x) \sqcup spherical(y) \sqcup ball(t)$). The part-of relation can be used to
represent lexical/ semantic relations, when it holds among nouns and
the kinds they denote (cf. Murphy 2010, Ch. 4; Asher 2011, Ch. 5).

The second proof system, the infomorphism, can shed light on how
polysemous nouns are related to concepts. In our account, concepts
also involve a form of "ambiguity". For instance, the *BALL* concept can
involve the union several visual types that can define its *SHAPE* range.
Our hypothesis predicts, however, that no type symmetry is necessary

between types of concepts and kinds. To see why this is the case, consider the mapping $g(ball)=ball'$ that can be established between nominal and visual types, with $ball'(x)$ a noun type. We have established that $ball'$ can be polysemous with respect to shape (i.e. $ball(x)$ can stand for $oval(x)$ or $spherical(x)$). Hence, the nominal type $ball'(x)$ can name the visual type $spherical(z)$ (i.e. $g(spherical(z))=ball'(x)$ holds).

Once this and the token mapping $f(x)=y$ are established, one can infer that a noun can describe a given object file. The indirect inference that the concept including this file and the corresponding kind are related can then be established (i.e. we have $i(ball)=BALL$). This mapping indirectly predicts that two types of asymmetry can arise. First, one can have a richer representation for noun types than for visual types. For instance, a speaker can use a potentially polysemous noun (here, *ball*) to describe an object with a specific function but not shape. Prototype effects may occur, too. Australian speakers may conceive *ball* as typically describing oval balls, because of the popularity of rugby and Australian Rules football codes; British speakers may prefer spherical balls because of football. However, rugby balls are often called *ovali* or *ovales* respectively in Italy and France; nouns *pallone* and *balon* are mostly reserved for Football (spherical balls). Our hypothesis predicts that such asymmetries can be systematic, once one looks beyond a given language to study the relation between the two systems via the lenses of our infomorphism.

6.3. *Relativism vs. Universalism*

The debate on the relation between language and cognition has witnessed a renewed interest over the three last decades. Recent literature suggests that strong versions of the hypothesis ("language determines cognition") but also strong Universalist positions ("cognition affects language") are untenable (Casasanto 2008, 2016). Here we (briefly) discuss how the infomorphism hypothesis is consistent with recent findings suggesting that an intermediate, "dynamic" relation between cognition and language is empirically adequate. Hence, we focus on works investigating this hypothesis in the domain of colour, *qua* a sub-domain of visual cognition that is partitioned via nouns or adjectives across languages (Berlin & Kay 1969; Kay & Maffi 1999).

Most works on the emergence of colour names investigate how speak-

ers of a given language can use words (lexical choices, in our terms) to distinguish possible non-linguistic concepts. Languages may partition the "colour" conceptual space according to a semantic/ lexical implicational hierarchy. If a language has a colour distinction at all, then it will distinguish between "black" and "white", or more accurately absence vs. presence of light. If the language distinguishes further colours, then it will distinguish between "red" and "green/ blue". Further distinctions can emerge, but always follow this hierarchy. These distinctions mirror basic visual mechanisms that organize humans' partition of the colour conceptual space (e.g. the distinction between *DARK* vs. *LIGHT* colour concepts).

Research on Papua New Guinean language Berinmo, however, found that this language creates a different partition than English when the yellow and green colours are involved. Speakers of this language have different terms from English for these colours, and thus label as "yellow" objects that would be labelled "green" in English (Roberson et al. 2002). This exception has been taken as initial evidence that the universal colour hierarchy may not be universal, after all. Subsequent works have investigated colour-naming across a much broader language sample (Regier et al. 2010; Regier et al. 2015). A general pattern emerging is that there seems to be a trade-off between polysemy, term specificity and communicative efficiency. Languages may have few terms to name colours, and these terms may be highly polysemous (e.g. Iduna, another Papuan language: Regier et al. 2015). Languages with richer inventories involve more efficient communication in context, but also a need to memorize a wider set of terms and the way they partition this semantic space (Kemp et al. 2018; Zaslavsky et al. 2018).

Our hypothesis is consistent with these results: it predicts that nominals for colours can be polysemous, and need not to name the same concepts across languages. An example of this latter prediction is the following. Across languages, the "blue" and "green" colours are sometimes named via the same nominal, one example being the Vietnamese *xanh* (Berlin & Kay 1969). In our approach, this nominal can be used to describe an object instantiating either a *BLUE* or a *GREEN* visual concept, or simply the join concept *BLUE⊔GREEN*. This naming relation can be represented via the mapping $g(xanh(x)) = green(x) \sqcup blue(x)$. What a Vietnamese speaker knows, according to our hypothesis, is not that

BLUE and *GREEN* are the same concept, but that this language lacks a distinct nominal for them, and that this nominal is ambiguous (i.e. polysemous) when used in context.

An English speaker, on the other hand, knows that *green* and *blue* are nominals (adjectives) that name the two distinct concepts. However, both types of speakers can develop *BLUE* and *GREEN* concepts, for these concepts represent visual distinctions independent of language. Our hypothesis can account for different naming patterns, relative to a language, because it models nouns as potentially labelling any concept combination. This fact can be captured by also assuming that *NS*, the nominal structure, is indexed with respect to a language. Thus, the infomorphism between $NS_{English}$ and *VS*, and $NS_{Vietnamese}$ and *VS* are not, and need not be, isomorphic. Languages can partition the conceptual space of colour, among others, in ways that can be modeled via our infomorphism, but that can vary in a subtle but logical manner. With this point in mind, we move to the conclusions.

7.   CONCLUSION

The analysis articulated in the preceding sections is a specific hypothesis about the relation between language and vision, or more precisely about the relation between the linguistic and the visual conceptualization of objects. We have seen the independent motivations for viewing the two, respectively, as maps of stimuli to visual features, and as lexically interpreted grammatical representations. We have seen how the two can be reinterpreted in terms of type-token maps, which in turn allows for a formalization of their interrelation as an infomorphism, with two functions defined over token and type variables from both domains. Finally, we have shown that the formal system thus defined proves, as necessary inferences, a number of facts about the relation between nominally– and visually–defined objects, fitting the empirical observation in a predictive way. Taking a step back, we conclude with some higher-level observations to place the analysis just summarized in its correct light.

The ever-present risk of surreptitiously identifying the structure of non-linguistic conceptual knowledge with the structure defined by the words of one particular language has been a constant concern, clearly

stated at the outset (Section 1). We have avoided the vicious circle of identifying mental content with the content of linguistic signs because the visual features (types) we use, labelled by capital words, correspond to natural language words just for convenience. The visual type represented as *blue* does not equal the content of the English word "blue", but is a symbol for a posited visual feature, understood as a property whose existence is justified by an account of vision. If it turns out that no such property is justified, then *blue* is no longer one of the visual features – with no consequence on the theory we have outlined, which is independent of the identity of each feature.

If language does not shape the basic ingredients of vision, the opposite is not true either. The discussion in Section 6.3 has shown why our results are incompatible with a strong Whorfian relativism; but that does not mean that language is irrelevant in shaping conceptual knowledge. Our position, in common with much work in the "constructionist" approach to grammar and lexical word classes, is that linguistic representations structure very basic conceptual ingredients, like the notions of individuality, boundedness, continuity, numerosity, divisibility, cumulativity, and, for verbal semantics, the notions of process, culmination, boundary, initiation, possibly also causation in some sense (see Ramchand 2011 for discussion). These, like other notions, often modeled by the help of mathematical tools, are presupposed ingredients in the linguistic construction of complex representations; but they do not "explain away" language, any more than they explain away vision. Specifically, the particular contribution of language lies in embedding and combining this basic information into symbols of arbitrary complexity (Spelke 2003; Acquaviva 2018).

Two final observations concern more specific aspects of our analysis. First, we distinguished sharply (Sections 5.2, 5.3) undefined from null values for the functions of the infomorphism: in the former there is no categorizing between a token and a type, in the latter there is, but with a (token or type) null value. This is not trivial, because it presupposes a specific perspective on the question of "empty" mental representations (see Rey 2005 for discussion). In our internalist approach, there is no problem in positing such empty mental representations, even on logically necessary grounds, as long as they have a function to perform – for instance representing the content of an unknown noun. Our

system thus makes a distinction between two senses in which a representation can be empty, which may be lumped together in a different perspective. Secondly, categorization and the classification of token elements as instances of a type has been the centre of our analysis. In our approach, categorization is fundamental not only because it underpins cognition *in* different domains (like vision), independently of language, but also because it plays a role in establishing the necessary connections between domains. Not only do we subscribe to the conclusion that "to cognize is to categorize" (Harnad 2005); insofar as our hypothesis about the language-vision relation is correct, we also offer a way to prove it.

ACKNOWLEDGEMENTS

Notes

[1]*Meet* and *join* are associative, idempotent and commutative operations. These aspects are not crucial to our discussion, so we do not discuss them further.

[2]Barwise & Seligman (1997, Ch. 2) presents this hypothesis as externalist-oriented, for they mostly analyze physical (i.e. mind-external) systems such as switchboxes. The internalist or externalist (or both) nature of the theory is thus reduced to its use: by applying this theory to mind-internal matters, we present an internalist application of the theory.

[3]Prototype theory includes a supplementary layer of complexity: a super-ordinate concept such as *BIRD* acts as the *PROTOTYPE* concept, in the absence of other concepts that can become preponderant in concept formation. The emergence of a prototype corresponds to the emergence of the pair of identities *ROBIN = PROTOTYPE* and *BIRD = PROTOTYPE*, i.e. to a subordinate concept being as prototypical as the super-ordinate concept. Nothing crucial hinges on this simplification.

References

Acquaviva, Paolo. 2018. 'Categorization as noun construction'. In Éric Mathieu, Myriam Dali & Zareikar Gita (eds.) 'Gender and Noun Classification', 41–63. Oxford University Press.

Acquaviva, Paolo & Panagiotidis, Phoevos. 2012. 'Lexical decomposition meets conceptual atomism'. *Lingue e linguaggio* 11: 165–180.

Aikhenvald, Alexandra. 2003. *Classifiers*. Cambridge University Press.

Asher, Nicholas. 2011. *Lexical Meaning in Context: A Web of Words*. Cambridge University Press.

Baker, Mark. 2003. *Lexical Categories*. Cambridge University Press.

Barwise, Jon & Seligman, Jerry. 1997. *Information Flow: The Logic of Distributed Systems*. Cambridge University Press.

Berlin, Brent & Kay, Paul. 1969. *Basic color terms: Their universality and evolution*. Univ of California Press.

Biederman, Irving. 1987. 'Recognition-by-components: a theory of human image understanding.' *Psychological review* 94, no. 2: 115–147.

Bisang, Walter. 2011. 'Word Classes'. In Jae Jung Song (ed.) 'The Oxford Handbook of Linguistic Typology,', 280–302. Oxford University Press.

Borer, Hagit. 2005a. *The normal course of events. Volume 2 of Structuring Sense*, vol. 2. Oxford University Press.

——. 2005b. *In name only. Volume 1 of Structuring Sense*. Oxford University Press.

Carey, Susan. 2009. *The origin of concepts*. Oxford University Press.

Casasanto, Daniel. 2008. 'Who's afraid of the big bad Whorf? Crosslinguistic differences in temporal language and thought'. *Language learning* 58: 63–79.

——. 2016. 'Linguistic relativity'. In Nick Riemer (ed.) 'Routledge handbook of semantics', 158–174. Routledge.

Casati, Roberto & Varzi, Achille. 1994. *Holes and Other Superficialities*. MIT Press.

Chierchia, Gennaro. 1998. 'Plurality of mass nouns and the notion of "semantic parameter"'. In Susan Rothstein (ed.) 'Events and grammar', 53–103. Kluwer.

Chomsky, Noam. 1995. *The minimalist program*. MIT press.

——. 2001. 'Derivation by phase'. In Michael Kenstowicz (ed.) 'Ken Hale: A life in language', 1–52. MIT Press.

Collins, John. 2017. 'The Semantics and Ontology of The Average American'. *Journal of Semantics* 34, no. 3: 373–405.

Cowan, Nelson. 2005. *Working Memory Capacity*. Psychology Press.

Croft, William. 1991. *Syntactic categories and grammatical relations: The cognitive organization of information*. University of Chicago Press.

Evans, Vyvyan. 2009. 'Semantic representation in LCCM Theory'. In Vyvyan Evans & Stéphanie Pourcel (eds.) 'New Directions in Cognitive Linguistics', 24, 27–56. John Benjamins.

——. 2010. 'Figurative language understanding in LCCM theory'. *Cognitive linguistics* 21, no. 4: 601–662.

——. 2015. 'A unified account of polysemy within LCCM Theory'. *Lingua* 157: 100–123.

Feldman, Jacob. 2007. 'Formation of visual "objects" in the early computation of spatial relations'. *Perception and Psychophysics* 69, no. 5: 816–827.

Fodor, Jerry A. 1975. *The Language of Thought*. Harvard University Press.

——. 1983. *Modularity of Mind*. MIT Press.

Ganter, Bernhard & Wille, Rudolf. 2005. *Formal concept analysis: mathematical foundations*. Springer.

Ganter, Bernhard, Stumme, Gerd & Wille, Rudolf. 2005. *Formal concept analysis: foundations and applications*, *Lecture Notes in Computer Science*, vol. 3626. Springer.

Gao, Tao, McCarthy, Gregory & Scholl, Brian J. 2010. 'The wolfpack effect: Perception of animacy irresistibly influences interactive behavior'. *Psychological science* 21, no. 12: 1845–1853.

Gärdenfors, Peter. 2000. *Conceptual spaces: The geometry of thought*. MIT press.

——. 2004. *Conceptual spaces: The geometry of thought*. MIT press.

Gibson, James Jerome. 1966. *The senses considered as perceptual systems*. Houghton Mifflin.

——. 1979. *The ecological approach to visual perception: classic edition*. Houghton Mifflin.

Gil, David. 2005. 'Isolating-monocategorial-associational language.' In Henrie Cohen & Claire Lefebvre (eds.) 'Categorization in cognitive science', 347–79. Elsevier.

Goldberg, Adele E. 1995. *Constructions: A construction grammar approach to argument structure*. University of Chicago Press.

——. 2006. *Constructions at work: The nature of generalization in language*. Oxford University Press.

Gupta, Anil. 1980. *The Logic of Common Nouns: An Investigation in Quantified Modal Logic*. Yale University Press.

Harley, Heidi. 2014. 'On the identity of roots'. *Theoretical linguistics* 40, no. 3-4: 225–276.

Harnad, Stevan. 2005. 'To cognize is to categorize: Cognition is categorization'. In Henrie Cohen & Claire Lefebvre (eds.) 'Handbook of categorization in cognitive science', 20–43. Elsevier.

Hauser, Marc D, Chomsky, Noam & Fitch, W Tecumseh. 2002. 'The faculty of language: what is it, who has it, and how did it evolve?' *Science* 298, no. 5598: 1569–1579.

Hummel, John E & Biederman, Irving. 1992. 'Dynamic binding in a neural network for shape recognition.' *Psychological review* 99, no. 3: 480–517.

Jackendoff, Ray. 2002. *Foundations of language: Brain, meaning, grammar, evolution*. Oxford University Press.

Jackendoff, Ray & Pinker, Steven. 2005. 'The nature of the language faculty and its implications for evolution of language (Reply to Fitch, Hauser, and Chomsky)'. *Cognition* 97, no. 2: 211–225.

Julesz, Bela. 1981. 'Textons, the elements of texture perception, and their interactions'. *Nature* 290, no. 5802: 91–97.

Kahneman, Daniel, Treisman, Anne & Gibbs, Brian J. 1992. 'The reviewing of object files: Object-specific integration of information'. *Cognitive psychology* 24, no. 2: 175–219.

Kay, Paul & Maffi, Luisa. 1999. 'Color appearance and the emergence and evolution of basic color lexicons'. *American anthropologist* 101, no. 4: 743–760.

Kearns, Kate. 2006. 'Lexical Semantics'. In Bas Aarts & April McMahon (eds.) 'Handbook of English Linguistics', 557–580. Blackwell.

Kelter, Stephanie & Kaup, Barbara. 2012. 'Conceptual knowledge, categorization, and meaning'. In Claudia Maienborn, Klaus von Heusinger & Paul Portner (eds.) 'Semantics: An International Handbook of Natural Language Meaning', 2775–2804. De Gruyter.

Kemp, Charles, Xu, Yang & Regier, Terry. 2018. 'Semantic typology and efficient commu-

nication'. *Annual Review of Linguistics* 4: 109–128.

Kirby, Simon. 2007. 'The evolution of language'. In Robin Dunbar & Louise Barrett (eds.) 'Oxford handbook of evolutionary psychology', 669–681. Oxford University Press.

Krifka, Manfred. 1995. 'Common nouns: A contrastive analysis of Chinese and English'. In Greg Carlson & Francis Jeffry Pelletier (eds.) 'The Generic Book', 398–411. University of Chicago Press.

Landau, Barbara, Smith, Linda B & Jones, Susan S. 1988. 'The importance of shape in early lexical learning'. *Cognitive development* 3, no. 3: 299–321.

Landau, Barbara, Smith, Linda B & Jones, Susan. 1992. 'Syntactic context and the shape bias in children's and adults' lexical learning'. *Journal of Memory and Language* 31, no. 6: 807–825.

Landman, Fred. 1991. *Structures for Semantics*. Kluwer.

Lasersohn, Peter. 2018. 'Common nouns as variables: Evidence from conservativity and the temperature paradox'. In Robert Truswell, Chris Cummins, Caroline Heycock, Brian Rabern & Hannah Rohde (eds.) 'Proceedings of Sinn und Bedeutung', vol. 21, 731–746.

Laurence, Stephen & Margolis, Eric. 1999. 'Concepts and cognitive science'. In Stephen Laurence & Eric Margolis (eds.) 'Concepts: Core Readings', 3–82. MIT Press.

Ledgeway, Adam. 2012. *From Latin to Romance: Morphosyntactic typology and change*. Oxford University Press.

Margolis, Eric & Laurence, Stephen. 2007. 'The Ontology of Concepts: Abstract Objects or Mental Representations?' *Noûs* 41, no. 4: 561–593.

Marr, David. 1982. 'Vision: A computational investigation into the human representation and processing of visual information'.

Massam, Diane. 2012. *Count and mass across languages*. Oxford University Press.

Michotte, Albert. 1963. *The Perception of Causality*. Methuen.

Mitroff, Stephen R, Scholl, Brian J & Wynn, Karen. 2005. 'The relationship between object files and conscious perception'. *Cognition* 96, no. 1: 67–92.

Murphy, Gregory L. 2002. *The big book of concepts*. The MIT Press.

Murphy, Lynn M. 2010. *Lexical Meaning*. Cambridge University Press.

Panagiotidis, Phoevos. 2014. *Categorial features*. Cambridge University Press.

Papafragou, Anna. 2005. 'Relations between language and thought: Individuation and the count/mass distinction'. In Henrie Cohen & Claire Lefebvre (eds.) 'Handbook of categorization in cognitive science', 245–277. Elsevier.

Pietroski, Paul, Lidz, Jeffrey, Hunter, Tim & Halberda, Justin. 2009. 'The meaning of 'most': Semantics, numerosity and psychology'. *Mind and Language* 24, no. 5: 554–585.

Pustejovsky, James. 1995. *The Generative Lexicon*. MIT Press.

——. 2013. 'Type theory and lexical decomposition'. In James Pustejovsky, Pierrette Bouillon, Hitoshi Ishara, Kyoko Kanzaki & Chungmin Lee (eds.) 'Advances in generative lexicon theory', 9–38. Springer.

Pylyshyn, Zenon. 1989. 'The role of location indexes in spatial perception: A sketch of the FINST spatial-index model'. *Cognition* 32, no. 1: 65–97.

——. 2004. 'Some puzzling findings in multiple object tracking: I. Tracking without keeping track of object identities'. *Visual cognition* 11, no. 7: 801–822.

Pylyshyn, Zenon W. 1984. *Computation and Cognition*. MIT Press.

Pylyshyn, Zenon W. 2006. 'Some puzzling findings in multiple object tracking (MOT): II.

Inhibition of moving nontargets'. *Visual Cognition* 14, no. 2: 175–198.

Ramchand, Gillian. 2011. 'Minimalist semantics'. In Cedric Boeckx (ed.) 'The Oxford handbook of linguistic minimalism', Oxford University Press.

Regier, Terry, Kay, Paul, Gilbert, Aubrey L & Ivry, Richard B. 2010. 'Which Side Are You on, Anyway?' In Barbara Malt & Phillip Wolff (eds.) 'Words and the Mind: How words capture human experience', 165–182. Oxford University Press.

Regier, Terry, Kemp, Charles & Kay, Paul. 2015. 'Word Meanings across Languages Support Efficient Communication'. In Brian MacWhinney & O'Grady William (eds.) 'The handbook of language emergence', 237–263. Wiley-Blackwell.

Rey, Georges. 2005. 'Philosophical analysis as cognitive psychology: The case of empty concepts'. In Henrie Cohen & Claire Lefebvre (eds.) 'Handbook of categorization in cognitive science', 71–89. Elsevier.

Riemer, Nick. 2010. *Introducing semantics*. Cambridge University Press.

Riesenhuber, Maximilian & Poggio, Tomaso. 2002. 'Neural mechanisms of object recognition'. *Current opinion in neurobiology* 12, no. 2: 162–168.

Rijkhoff, Jan. 2002. *The noun phrase*. Oxford University Press.

Roberson, Debi, Davies, Ian & Davidoff, Jules B. 2002. 'Color categories are not universal: Replications and new evidence'. In Barbara Saunders & Van Jaap Brakel (eds.) 'Theories, technologies, instrumentalities of color: Anthropological and historiographic perspectives', University Press of America.

Rosch, Eleanor. 1973. 'Natural categories'. *Cognitive psychology* 4, no. 3: 328–350.

——. 1975. 'Cognitive representations of semantic categories.' *Journal of experimental psychology: General* 104, no. 3: 192.

——. 1978. 'Principles of Categorization'. In Eleanor Rosch & Barbara B. Lloyd (eds.) 'Cognition and Categorization', 27–48. Lawrence Erlbaum Associates.

Sadock, Jerrold M. 2012. *The modular architecture of grammar*. Cambridge university press.

Sag, Ivan A, Boas, Hans C & Kay, Paul. 2012. 'Sign-based construction grammar: An informal synopsis'. In Hans C Boas & Ivan A Sag (eds.) 'Sign-based construction grammar', 69–202. CSLI Publications.

Scholl, Brian J. 2001. 'Objects and attention: The state of the art'. *Cognition* 80, no. 1-2: 1–46.

——. 2009. 'What have we learned about attention from multiple object tracking (and vice versa)'. In Don Dedrick & Lana Trick (eds.) 'Computation, cognition, and Pylyshyn', 49–78. MIT Press.

Scholl, Brian J & Gao, Tao. 2013. 'Perceiving animacy and intentionality: Visual processing or higher-level judgment'. In M.D. Rutherford & Valerie A Kuhlmeier (eds.) 'Social perception: Detection and interpretation of animacy, agency, and intention', 197–230. MIT Press.

Scholl, Brian J & Pylyshyn, Zenon W. 1999. 'Tracking multiple items through occlusion: Clues to visual objecthood'. *Cognitive psychology* 38, no. 2: 259–290.

Scholl, Brian J, Pylyshyn, Zenon W & Feldman, Jacob. 2001. 'What is a visual object? Evidence from target merging in multiple object tracking'. *Cognition* 80, no. 1-2: 159–177.

Senft, Gunter. 2000. *Systems of nominal classification*. Cambridge University Press.

Serre, T, Wolf, L & Poggio, T. 2005. 'Object recognition with features inspired by visual cortex'. In 'Proceedings of 2005 IEEE Computer Society Conference on Computer

Vision and Pattern Recognition (CVPR)', vol. 2, 994–1000. IEEE Computer Society Press.

Smith, Barry. 1995. 'Formal ontology, common sense and cognitive science'. *International journal of human-computer studies* 43, no. 5-6: 641–666.

——. 2003. 'Ontology (Science).' In Carola Eschenbach & Michael Grueninger (eds.) 'Formal Ontology in Information Systems', 21–35. IOS Press.

Spelke, Elizabeth S. 2003. 'What makes us smart? Core knowledge and natural language'. In Dedre Gentner & Susan Goldin-Meadow (eds.) 'Language in mind: Advances in the study of language and thought', 277–311. MIT Press.

Talmy, Leonard. 2000. *Toward a cognitive semantics*. MIT Press.

Tomasello, Michael. 2003. *Constructing a language: A usage-based theory of language acquisition*. Harvard University Press.

Tyler, Andrea & Evans, Vyvyan. 2003. *The semantics of English prepositions: Spatial scenes, embodied meaning, and cognition*. Cambridge University Press.

Ullman, Shimon. 1979. *The interpretation of visual motion*. MIT Press.

——. 2000. *High-level vision*. MIT Press.

Van Eijck, Jan & Zwarts, Joost. 2004. 'Formal concept analysis and prototypes'. In Werner Kuhn, Martin Raubal, Florian Probst & Krysztof Jonowicz (eds.) 'Proceedings of the Workshop on the Potential of Cognitive Semantics for Ontologies', 1–7.

Wells, Andrew J. 2002. 'Gibson's affordances and Turing's theory of computation'. *Ecological Psychology* 14, no. 3: 140–180.

Wierzbicka, Anna. 1988. *The semantics of grammar*. John Benjamins Publishing.

Withagen, Rob, De Poel, Harjo J, Araújo, Duarte & Pepping, Gert-Jan. 2012. 'Affordances can invite behavior: Reconsidering the relationship between affordances and agency'. *New Ideas in Psychology* 30, no. 2: 250–258.

Zacks, Jeffrey M & Swallow, Khena M. 2007. 'Event segmentation'. *Current directions in psychological science* 16, no. 2: 80–84.

Zamparelli, Rob. 2000. *Layers in the determiner phrase*. Garland.

Zaslavsky, Noga, Kemp, Charles, Regier, Terry & Tishby, Naftali. 2018. 'Efficient compression in color naming and its evolution'. *Proceedings of the National Academy of Sciences* 115, no. 31: 7937–7942.

Zhu, Song-Chun, Guo, Cheng-En, Wang, Yizhou & Xu, Zijian. 2005. 'What are textons?' *International Journal of Computer Vision* 62, no. 1-2: 121–143.