Department of Computer Science,
University of Otago

Technical Report OUCS-2017-03

A model of object property representations: visual object classification, working memory and the syntax of predication

Authors:

Alistair Knott, Lech Szymanski, Brendan McCane
Department of Computer Science, University of Otago, New Zealand

Martin Takac
Centre for Cognitive Science, Comenius University, Slovakia

Department of Computer Science,
University of Otago, PO Box 56, Dunedin, Otago, New Zealand

http://www.cs.otago.ac.nz/research/techreports.php

A model of object property representations: visual object classification, working memory and the syntax of predication

Ali, Lech, Brendan, Martin

February 12, 2016

1 Introduction

Say an observer looks at a dog, and reports one of its properties that happens to be salient: *Hey, this dog is dirty!* (Or alternatively *This dog is [hairy!/white!/small!/angry!]* and so on.) This process involves a mixture of perceptual and linguistic mechanisms.

On the perceptual side, the observer must identify a specific object in the world, classify the object as a dog, and additionally identify a particular property of the object, such as 'dirty' or *hairy*. These mechanisms are most often studied with reference to *visual* perception, where the focus is on **visual attention** and **visual object classification**. There is a large literature on these topics, both in experimental neuroscience () and in computer modelling (). There is a large literature on how object categories are represented in the brain (), and there are interesting recent results on the neural areas holding representations of visual properties relating to shape, colour and texture (). But the mechanism by which an observer singles out an individal property of an attended object is not yet understood.

On the linguistic side, the observer must produce a *sentence* that reports the property, in English or in some other language. This sentence is a **predicative** sentence. The cognitive representations of sentences and their structure are studied within the field of linguistics, and in particular, theoretical syntax. In this field, the syntactic structure of predicative sentences is particularly controversial. There are many alternative models of predicative sentences, and as the topic of predication connects with a number of other controversial topics, finding an adequate model is extremely challenging.

Research on visual object perception makes very little reference to theoretical syntax, and the reverse is also true. In this paper, we argue that these areas of research can be helpful to one other, particularly in the resolution of recalcitrant or controversial questions. Specifically, we argue that ideas from syntax suggest useful models of property perception, and that models of vision help to resolve controversies about predicate structures in linguistics. The assumption underlying this argument is that *linguistic representations make reference to perceptual processes*. There is in fact a reasonable amount of evidence for this hypothesis, and it is gaining some prominence in cognitive science (though not without some controversy: see e.g. ??). However, the hypothesis has not yet affected *methodologies* in linguistics or perceptual neuroscience: research in these disciplines proceeds as if there were no connection between them. In the current paper, we propose a novel model of predication that is partly *motivated* by a model of visual perception. We argue this model addresses open questions in linguistics and in visual neuroscience—and more far-reachingly, that its success provides evidence that these topics should be studied together, rather than independently.

We begin in Section 2 and **??** by reviewing current models of visual property perception and predicative sentence structure. In Section 4 we introduce the framework for our own account: a neural network model of semantic working memory representations, which interfaces both with the perceptual system and with language. In Section 5, we introduce a neural network model of property perception, which supervenes on a standard model of visual object classification. In Section 6, we introduce our proposal about how this model interfaces with a model of syntactic representations, and ultimately with an account of sentence production.

2 Visual property perception

On the perceptual side, the observer must identify a specific object in the world, classify the object as a dog, and additionally identify a particular property of the object, such as 'dirty' or *hairy*. These mechanisms are most often studied with reference to *visual* perception, where the focus is on **visual attention** and **visual object classification**. There is a large literature on these topics, both in experimental and brain-imaging paradigms () and in computer modelling (). Until recently, not much attention has been paid to the question of how the visual system identifies *properties*. The perceptual mechanism that identifies properties is presumably closely related to the mechanism that identifies categories, but the precise relationship is still unclear.

2.1 Representations of object types

Representations of general concepts are often said to reside in **semantic memory**, which is distinguished from episodic memory. Semantic memory is defined differently by different people: for Binder and Desai (2011) it includes 'all the declarative knowledge we acquire about the world' and includes 'the names and physical attributes of all objects, the origin and history of objects, the names and attributes of actions, all abstract concepts and their names, knowledge of how people behave and why, opinions and beliefs, knowledge of historical events, knowledge of causes and effects, associations between concepts, categories and their bases'.

[1]

Worth mentioning Macoir *et al.* (2015): they find that patients with progressive semantic dementia have relatively spared performance for adjectives. (This supports a model where nouns are read from convergence zones, while adjectives are read from media contributing to these convergence zones.)

You should also mention theories about the structure of object type concepts: e.g. McCarthy and Warrington's (1988) idea that the important properties of living things relate to sensory properties (shape, colour etc) while the important properties of nonliving things relate to functional properties. Also Patterson *et al.*'s (2007) 'semantic hub' model, which sees a modality-independent convergence zone in anterior temporal cortex holding object type representations. Which sort of synchs up with Binder and Desai's (2011) view (where a large part of temporal cortex, plus inferior parietal cortex, is involved in holding modality-unspecific convergence zones).

Also worth mentioning Pulvermüller and Hauk's (2006) study of 'colour and form words' and their separate neural substrates (). [They don't give the list, or even specify if they're nouns or verbs: I think they're a mixture of both.]

Coutanche and Thompson-Schill (2015) have evidence that activity in anterior temporal cortex can be used to decode representations of object types even when the object in question is not perceptually present, but merely *anticipated*. (To me this suggests a copy in temporal cortex of a pattern stored prefrontally.)

2.2 Neural representations of colour and colour adjectives

There are two brain regions that seem to be particularly activated by colour (though neither of them *only* represent colour): an area in the lingual gyrus labelled V4, or sometimes VO1, and an area in the medial fusiform gyrus, labelled V4α (see Murphey *et al.*, 2008 for a review). Murphey *et al.* tested a human patient with electrodes implanted in the latter region: these responded more to chromatic than non-chromatic stimuli, and were selective for colour, particularly blue/purple; stimulation of these electrodes in the absence of visual stimuli elicited the percept of a blue/purple colour near the fovea.

There's evidence that colour percepts activate V4α, even when induced by a colourless stimulus (Morita *et al.*, 2004). Coutanche and Thompson-Schill (2015) found that the colour of a visually presented stimulus

[1] We identify another type of closely-related type of memory, **token object memory**, which holds information about the properties of particular objects. Thus the fact that my dog Fido is brown resides in token object memory, while the fact that dogs are often brown resides in semantic memory. The two types of memory are closely related because the generalisations in semantic memory are learned from the facts in token object memory. We will group these two types of memory together into a type of memory called **stative memory**, and distinguish stative memory from episodic memory.

was represented in anticipatory activity in right V4, even prior to a stimulus being presented. (Shape was also represented in V4, in fact).

2.3 Neural representations of texture

Cant and Goodale (2007) showed subjects nonsense objects that varied in form and in surface properties (naturalistic textures like marble or wood grain, in different colours). While attention to form activated the lateral-occipital (LO) area, as expected, attention to texture (and to a lesser extent, colour) activated two distinct regions, the collateral sulcus and the inferior occipital gyrus. This was replicated in a fMRI adaptation study (Cant *et al.*, 2009). (Neither study found extrastriate regions sensitive only to colour.) Cavina-Pratesi *et al.* (2010), using a fMRI technique examining selective rebound from adaptation, found distinct areas sensitive to form (LO), colour (anterior contralateral sulcus and lingual gyrus), and texture (posterior contralateral sulcus). This texture region is different to those found by Cant *et al.*, perhaps because the textures in the experimental stimuli were tactile textures (rough, smooth, spiky etc). (All Cant *et al.*'s stimuli were smooth and required vision to be identified.) The areas identified by Cavina-Pratesi *et al.* were also consistent with data from patients with selective deficits in identifying colours, shapes and forms.

Cavina-Pratesi *et al.* (2010) also found regions in the fusiform gyrus that were selective to specific combinations of shape, colour and texture.[2]

2.4 Neural representations of shape and shape adjectives

Coutanche and Thompson-Schill (2015) found that the shape of a visually presented stimulus could be decoded from right V4 (but not its colour).

The most common shape adjectives are relative ones, like things like *big*, *thin*.

2.5 Neural representations of category-relative properties

Cite Chen *et al.* (2014).

There's evidence that size adjectives aren't automatically generated for object types, except for types on the extremes of absolute size (elephant, mouse). Chen *et al.* argue that size adjectives are only generated in response to a query. (I think they're generated automatically for token objects, if the object's size is sufficiently different from the expected size of objects of its type.)

2.6 Neural representations of affective properties

Emotive perceptual stimuli (including objects) activate a network of brain regions, including the amygdala for fear (Öhman, 2005), the anterior insula for disgust (Wicker *et al.*, 2003), the ventral occipital cortex for facial attractiveness (Chatterjee *et al.*, 2009) and the medial orbito-frontal cortex for aesthetic judgements (Ishizu and Zeki, 2011).

There is a certain amount of evidence for an autonomous system processing the emotional valence of perceived stimuli, though exactly what this processes must be carefully delineated (see Pessoa, 2005) for a good review). There's quite good evidence that at least some forms of perceptual processing of emotional content require focal attention; but there is also evidence that some aspects of emotional content are processed pre-attentionally. (This makes sense if you consider that emotional content contributes to a computation of salience, that might trigger an interruption of the agent's current task, à la Corbetta and Shulman.) There's also evidence that while some emotional processing is early, other more sophisticated processing is late: again see Pessoa (2005) for a discussion. With these caveats, here is a discussion of the autonomous system for processing emotional properties of objects.

[2]The fusiform gyrus runs right along temporal cortex, above the inferior temporal gyrus and below the parahippocampal gyrus, but it would be generally regarded as 'inferior temporal' (IT). Neurons in monkey IT cortex are well known to be sensitive to complex stimuli, including combinations of colour, form and surface pattern (Komatsu and Ideura, 1993).

Fear and the amygdala There are two routes to the amygdala during perceptual experience: a 'fast' route through subcortical regions (the superior colliculus and pulvinar), and a 'slow' route via temporal cortex (see Öhman, 2005). Fear-eliciting stimuli activate the amygdala prior to visual cortex (ses again Öhman, 2005).

Attractiveness and the ventral occipital cortex The attractiveness of human faces is also recognised very fast. Olson and Marshuetz (2005) presented subjects with pictures of attractive and unattractive faces at very short exposures: short enough that they reported they could not see a face at all. (Stimuli were masked to eliminate persistence of vision.) Subjects' judgements of facial attractiveness was nonetheless significantly better than chance. There are also areas of the visual pathway that appear to be automatically activated by facial attractiveness, in particular the 'ventral occipital cortex', which includes the fusiform face area and the lateral occipital cortex (see Chatterjee *et al.*, 2009).

Aesthetic properties and orbitofrontal cortex There is some evidence that a domain-independent representation of 'aesthetic beauty' is activated in medial orbito-frontal cortex. For instance, Ishizu and Zeki (2011) found activity in this area for both pictures judged beautiful and excerpts of music judged beautiful, and the activity in this area was proportional to the judged degree of beauty. (Orbitofrontal cortex is also differentially activated by stimuli judged beautiful and ugly; see Kawabata and Zeki, 2004.) Again there seem to be both fast and slow responses in this area: some judgements of beauty are considered, while others are fast (though there's no indication at all that this area generates evaluations that arrive faster than neutral class labels).

Look-ahead to linguistic/syntactic issues There's evidence that reading emotional adjectives activates the (left) amygdala more than reading neutral adjectives—and that positive adjectives elicit more activity than negative ones (Herbert *et al.*, 2009). Emotional adjectives also activate the (left) inferior/middle occipital gyrus (BA18/19), and superior frontal gyrus (BA9). So some of the same areas that are activated by emotion-eliciting stimuli.

The emotive responses that are *post* focal attention, but *pre* classification, are particularly interesting, because that's where they sit in the syntactic structure, according to our general hypothesis.

3 The syntax of adjectives and predication

On the linguistic side, the focus is on the sentence that reports the property (in English or in some other language). This sentence is a **predicative** sentence: its subject is a referential determiner phrase (DP),[3] and its predicate is an adjective (specifically, a **predicative adjective**). In English, and many other languages, the sentence's main verb is the **copula**, *to be*; however, there are also many languages where the relation between subject and predicate is unmarked by any explicit word, for instance, Chinese and ?? (). The syntax of predicative sentences, and of copular sentences in particular, is extremely controversial. For one thing, copular sentences have functions beyond predication. They are also used to assert statements of 'token identity' between two referents (e.g. *The morning star is the evening star*). And there are a variety of special predicates. Some of these relate to the location of an object, rather than its intrinsic properties: these are often reported with different syntactic structures (). Other predicates relate to a role that the object plays (e.g. *X is the winner* or *X is the president*); these also sanction alternations on the standard template for copular sentences (). Moreover, for completeness, a syntactic account of predicative adjectives demands an account of **referential adjectives**: that is, adjectives appearing inside referential DPs, such as *the dirty dog*. The syntax of DP-internal adjectives is a hugely controversial topic in its own right.

3.1 Inverse copulas and token-identity statements

There are some predicative sentences which can be reversed. For instance:

[3] We use this term in preference to 'noun phrase' (NP), as is now conventional in linguistics (see e.g. Abney, 1987).

(1) John was the winner.

(2) The winner was John.

Or more elaborately:

(3) The picture of the wall was the cause of the riot.

(4) The cause of the riot was the picture of the wall.

These may seem to be symmetrical, but they're not, as Moro (1997) has shown using extraction phenomena:

(5) Which riot was the picture of the wall the cause of?

(6) *Which wall was the cause of the riot the picture of?

This suggests that *the picture of the wall* is the subject and *the cause of the riot* is the predicate, regardless of which way round these sentences appear; likewise, *John* is the subject and *the winner* is the predicate in both cases. Moro suggests a scheme whereby the copula has an empty subject, and introduces a predicative small clause as its complement, featuring the true subject and the predicate, one of which must raise into the higher subject position. Heycock and Kroch (1997) modify this proposal, so that it applies when the copula is used as an **equative**, that asserts identity between two things. On their analysis, 'the cause of the riot' and 'the president' are special in being predicates that apply to a single referent, and therefore can be used referentially. When used referentially, they refer to an object by its *role*, rather than by its intrinsic properties.

3.2 Predication in VSO languages

In VSO languages, the predicate is also fronted. For instance in Niuean:

(7) Ko e kamuta a au
TAM art carpenter art I
'I am a carpenter'

Linguists often try to explain this fronting in terms of the mechanism that raises verbs to a high position (e.g. Carnie, 1995; Massam and Smallwood, 1997). However, this introduces some complications, because the element being raised is a maximal projection, rather than a head. (Note the fronting of the predicate here is different from that found in equative copular sentences: it happens for truly predicative sentences, as well as for equatives.) If there's some SM principle that can be seen as the basis for predicate fronting, that would be nice.

3.3 DP-internal adjectives

Predicative and attributive adjectives There's a well-known distinction between **predicative** adjectives, that are expressed in the complement of a copula clause, and **attributive** adjectives, that modify a noun in a DP. Actually some DP-internal adjective phrases are better classed as predicative: namely, those which can be analysed as predicates of a post-nominal reduced relative clause, such as this one:

(8) A boy [$_{AP}$ proud of his father]

(Which is best analysed as *A boy [$_{CP}$ who is [$_{AP}$ proud of his father]]*, with elided relative pronoun and copula; see Alexiadou *et al.*, 2007.)

Prenominal adjectives In English and many other languages, *prenominal* adjectives cannot take complements.

(9) *The [proud of Mary] boy

(10) *The [happy that it's sunny] boy

But this isn't a universal principle: prenominal adjectives with complements are possible in several languages, including German, Greek and Swedish (see Cabredo-Hofherr, 2010). Here's an example from German:

(11) Die auf ihren Sohn stoltze Mutter
The of-her-son-proud mother
'The mother who is proud of her son'

Interpretive differences between pre- and post-nominal adjectives Bolinger (1967) noticed that there's a semantic difference between *the navigable rivers* and *the rivers [that are] navigable*: the latter expression denotes the rivers that are *currently* navigable, while the former denotes the rivers that are 'enduringly' navigable.

Intensional/non-intersective adjectives Some adjectives don't simply apply predicates to the denotation of the noun, but modify the sense of the noun. Examples include adjectives like *present* and *former*. These are termed **intensional** or **non-intersective** adjectives.

(12) The former president

(13) ?? The president is former.

Some other examples are *alleged* (as in *an alleged murderer*) and *fake* (as in *a fake gun*).

A nice idea: *former* and *present* could be operations that temporarily change the reference time (i.e. the situation), and therefore assign a different type to the individual referred to. (It's important this change is only temporary.) *Alleged* could work this way too, if situations also represent belief states (as they naturally do in our model, though we haven't worked it out). And *fake* can work like this, if we assume the situation in question is the one in which the fake object *passes for* the genuine object.

The important question is what would *prompt* a temporary change in situation. Some other situation must compete to be adopted, in the same way it does when an agent is reminded of something.

Nouns that denote roles, and deverbal nouns Evaluative adjectives like *good* and *bad* are also a type of non-intersective adjective. They're not quite the same as the ones discussed above: a good guitarist is a guitarist, and even a hopeless guitarist is in some sense a guitarist. However, *good* is certainly not a property that's predicated of the individual denoted by *guitarist*, so it has this in common with *former*, *fake* etc. Instead, *good* refers to the way the individual *fulfils the role* of guitarist. (He's good at playing the guitar, or at being a teacher, or CEO, or husband.)

A role is something active, it's something you do. We've already seen some special nouns that refer to roles, like *president* and *teacher*, in the context where a unique individual plays a role (see the above discussion on inversion in copula sentences). There's a whole class of nouns that identify individuals in roles, namely the class of **deverbal nouns** like *dancer* and *actor*. (In the general case, these roles are not necessarily unique: there has to be one president, and in a given situation one teacher, but thewe are few situations where there must be just one actor or one dancer.

When I say *Sue is a dancer*, I am saying she has, or can take on, the role of dancer. How should this role be represented in our system? One way to start is by considering how an observer would directly *recognise* that Sue is a dancer. He would do this by noticing Sue dancing. But clearly he has to use this observation *to update his representation of Sue*, rather than just to add an episode to episodic memory. His representation of Sue is stored in semantic memory, which in our scheme is a system of associations between LTM individuals and properties. In this case, the properties relate to *actions* that Sue does.

One possibility is that the observer doesn't recognise a completed event in which Sue dances, but rather recognises that *Sue is dancing*. I've already suggested a semantics for the progressive clause *Sue is dancing*, in which there are direct associative links between LTM individuals and LTM episodes in which they are the agent. If the observer notices Sue dancing regularly, associations will develop between the LTM individual representing Sue and the generic LTM episode unit 'person dances'. This unit abstracts over token times, since the dancing happens on multiple occasions; it also abstracts over token agents, because 'person dances' is encountered much more frequently than 'Sue dances'. Dancers also do certain other actions with high frequency; for instance they do stretching exercises, go to dance shows, give dance lessons, drink smoothies. Whenever we activate the LTM individual ‘Sue’, we activate a whole *set* of generic episodes of the form ‘person Xs’. And the same thing happens when the observer activates other LTM individuals representing dancers, who have similar patterns of activity. The key idea is that these generic episodes *provide input to the system that learns object types*: so the observer learns a type associated with ‘dancer’. Of course this system also takes input from regular physical properties, so the facts that dancers wear scruffy clothes, leg-warmers, are thin, supple etc also contribute to the definition of the ‘dancer’ type.[4][5]

Now consider again evaluative adjectives modifying deverbal nouns. These are very often ambiguous, as in this case:

(14) Sue is a beautiful dancer

This can either mean ‘Sue is a dancer, and Sue is beautiful’, or ‘Sue is a dancer, and Sue dances beautifully’. Larson (1998) suggests that a deverbal noun has two argument positions, one for an individual, and one for an event, and that either can be modified by a prenominal adjective. Let’s say a deverbal noun denotes an object type that’s partly defined by associations with generic LTM episodes: for instance *dancer* denotes the type ‘dancer’, which is defined partly by association with the generic LTM episode ‘person dances’. This type is certainly associated both with an individual (the one it’s predicated of) and an episode (‘person dances’). So there’s scope for an implementation of Larson’s idea.

Now we have to consider how an evaluative adjective like *good* or *beautiful* can apply to a person’s fulfilment of a role. One idea is to assume there is an ideal notion of what actions the role should include (e.g. for a president), represented as a distribution over episodes—then a person’s own distribution of episodes can be measured against that. Another idea is that there is an ideal notion of how individual actions should be *performed*, and a person’s own performance can be measured against that. I hope these two things are the same, at some level. These ideals are cultural things: they’re taught partly by people modelling the roles, and partly by people making statements about them.

Aside An interesting study by Kemmerer *et al.* (2012) suggests that patients with brain lesions can show specific impairment in their knowledge of adjective-ordering principles (e.g. not be able to identify *blue thick towel* as anomalous), while being unimpaired in their ability to spot ungrammatical adjective-noun combinations (e.g. *towel blue*). The areas most commonly associated with this pattern were the left posterior inferior frontal gyrus and the left inferior parietal lobule.

3.4 Head movement in the DP

N-to-D raising in Danish In Danish, there’s evidence that N can raise right up to D. In one paradigm, determiners appear before nouns, as in English:

(15) det gamle hus
the old house

But there’s an alternative form where definiteness is expressed as an inflection of the noun:

(16) hus-et
house the
‘the house’

One tricky thing for me is that the N can’t raise to D if there’s an intervening adjective.

[4] In fact, maybe it’s not LTM episodes as such that are associated with LTM individuals, but *situations*, which encode the occurrence of particular episodes *at particular times*. A whole set of situations could be activated by a LTM individual in parallel. I like this idea. But I’ll keep talking about sets of LTM episodes for now.

[5] With this premise, we can give a better definition of the circumstances in which an observer’s attention is drawn to an agent’s participation in a particular episode, of the kind that results in a progressive like *Sue is eating a hamburger*. Let’s say we have classified Sue as a dancer, using a mixture of actions and physical properties. As we watch her, we activate a distribution over expected actions, and then subtract that: if what she’s doing right now is expected, it’s not of interest, but it it’s unexpected (dancers don’t eat hamburgers!), we record it as a stative property.

(17) * hus-et gamle

Delsing (1993) explains this by positing that prenominal adjectives are actually heads: since these don't raise, their presence blocks movement of N to D. But even if you accept the head movement constraint (which I don't think I do), there are other languages where definitess is signalled on a high noun in the presence of an adjective, for instance Romanian (Giusti, 1997).

(18) acest frumos baiat
this nice boy

(19) baiat-ul acesta frumos
boy-definite this nice
'this nice boy'

I think we have to explain the ungrammaticality of a high N with adjectives in Danish as a language-specific convention. (Our network can certainly learn such conventions, as well as other ones like the Romanian construction.)

Snowballing and reversed DP structures In some African languages, Gungbe and Fongbe, the surface word order in DPs looks completely reversed. Here's an example from Gungbe:

(20) távò xóxó dàxó ló
table old big the
'the big old table'

Aboh (2000) assumes a standard right-branching structure for DPs (DP introducing a functional projection FP for *big* then NP, which hosts the adjective *old*): the reversed surface order is derived by assuming 'snowballing', whereby when the N head raises to D, it *collects* the material in the successive XPs it passes through. In this analysis, the constituent raises through specifier positions. Thus what raises from NP to [Spec, FP] is not just *table* but *table old*. In this position it collects the specifier *big*, so what raises to [Spec, DP] is *table old big*. In this position it collects the D head, and we have *table old big the*. Hmm.

As far as I can tell, in my model, the only way of achieving this is to allow prepared sequences to be rehearsed *in reverse*. There's certainly precedent for that in neuroscience, though I'm not sure that people can *overtly* rehearse prepared SM sequences in reverse with any facility. I'm pretty sure they can't. So this would have to be a model of something that happens below the level of conscious awareness. (We're certainly not aware of the SM sequences I'm talking about in this model.)

3.5 A mixed model of DP-internal adjectives

To summarise all this: we assume a model where some adjectives originate as specifiers of projections in the DP, along the lines suggested by Cinque (1994). The XPs in a DP are as follows: DP, NumP, EvalP, Size/ShapeP, NP. Evaluative adjectives (e.g. *beautiful*) appear in [Spec,EvalP]; Size/shape adjectives (e.g. *big*, *tall*) appear in [Spec,Size/ShapeP]; regular attributive adjectives appear in [Spec, NP]. A special projection 'NonIntP' can optionally appear either after NumP or after EvalP. Specifiers of this projection hold nonintersective adjectives like *former* and *alleged*.

Nouns, and other heads, can appear at any of the head positions, but there are certain positions that are typical. In English and Germanic languages, nouns appear at the lowest head position (the head of NP). In French, nouns normally appear at the head of Size/ShapeP, so they appear after all adjectives except the attributive ones. In languages like Romanian and Danish, nouns can appear on the head of DP, combined with definiteness.[6]

Other adjectives originate as predicates in a reduced relative clause introduced as the complement of NP. I'll call these **clausal adjectives**. These appear postnominally in nearly all languages (for instance in English and Romance languages). In Romance languages, there are two kinds of postnominal adjectives: those generated in the specifier of DP (which are postnominal due to N being pronounced at a higher head),

[6]This may happen in Italian in *Giovanni mio* too; see Longobardi, 1994.

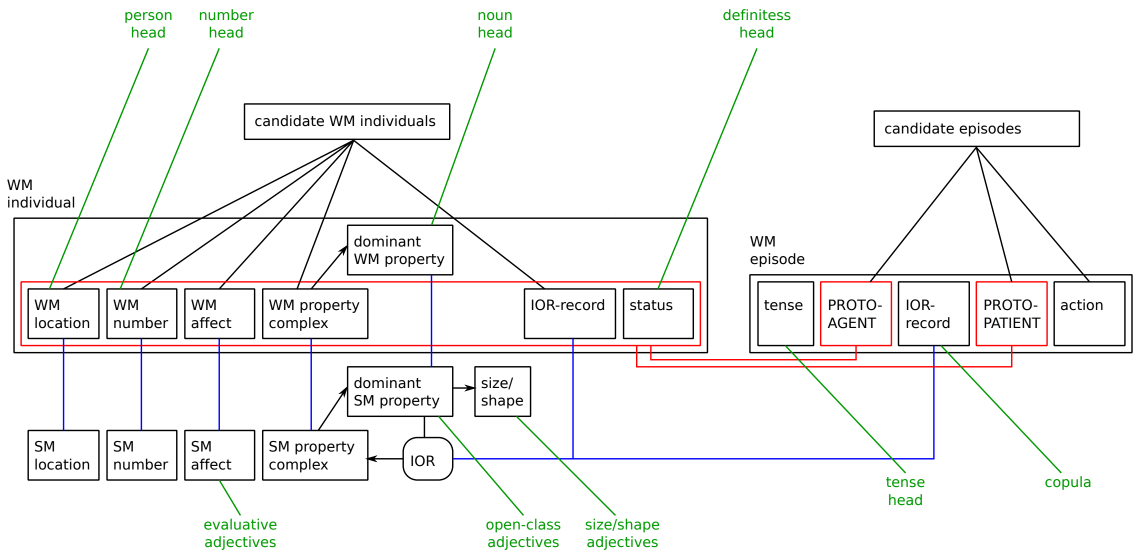

Figure 1: Architecture of the WM and SM networks, with interfaces to surface language

and those generated as part of a reduced relative clause. These latter ones come strictly after the specifier-generated ones. I want to use these clausal adjectives to explain now only English examples like *A man [proud of himself]*, but also French examples which appear to violate rules about adjective ordering, such as the following:

(21) une énorme maison magnifique
an enormous house magnificent

I'll suggest that *magnifique* is part of a relative clause, and thus doesn't appear within the hierarchy of projections that explain cross-linguistic regularities in adjective order.

There are a few languages which allow clausal adjective phrases to appear prenominally: see for instance the German example 11 ('the of-her-son-proud mother'). We'll explain these with a hack, using the word-sequencing network. Say this network takes as an input the fact that a relative clause will be generated, at the time when the head noun ('mother') is to be pronounced. It can choose to delay pronunciation of the head noun, recording the fact that it's needed in its hidden state, then pronounce the reduced relative clause, and then produce the head noun. Likewise in Greek, and a few other languages. We don't need to hypothesise any mechanism in 'LF' to account for this structure: we can just say it arose culturally, perhaps through some meta-linguistic game, and became conventionalised.

4 A neural network model of the language-perception interface

5 A neural network model of property perception

6 A neural network model of predicative and referential adjectives

6.1 Ideas about IOR and noun/adjective syntax

Here's an idea:

- There's an RPC medium and a type medium in the SM system: these media are active transiently during rehearsal of a WM individual.

- There's an RPC medium and a type medium in the WM individuals system: these media are active tonically during rehearsal of a WM individual.
- Nouns are read from the WM type medium; adjectives are read from the SM type medium.
- The difference between DP-internal adjectives and **predicate adjectives** (i.e. adjectives appearing as the predicate in a predicative clause) is that:
  - A DP with an adjective in it is read from a WM individual that encodes an object-establishing SM sequence *with one extra SM operation in it*: IOR. Importantly, while the IOR operation is *planned* in the WM individual, it's only *executed* in the SM system, so it generates a property like 'black' as a side-effect in the *SM* 'type' medium, while the original dominant property assembly (DPA) 'dog' is still active in the WM medium.[7] This means that you can read out *dog* before *black* (e.g. by reading out the head of NP then its spec), or you can read out *black* before *dog* (by reading out the spec of NP, and then the head of IORP.)
  - A predicative clause is read out from a different WM structure. Here, (i) You establish an object without any IOR, and *create a regular WM individual* representing this object (e.g. *the dog* [unmodified]); (ii) You do IOR; (iii) You create *another* WM individual representing the modified object representation. The two individuals are referred to in the AGENT and PATIENT fields of the WM episode, but these references are ambiguous, because they point to the same object: so in reading out the clause, there are two possible orders (subject predicate, and predicate, subject) and the one you pick is a matter of linguistic convention.

    IOR still happens in the SM system in this case, but crucially, when creating a WM individual representing the results of IOR, the modified SM RPC and DPA media are *copied back into the WM RPC and DPA*. The resulting WM individual could end up with an object category in the DPA, which is most readily pronounced as a noun (and we get a predicate nominal). Or it would end up with a property like 'black' in the DPA: in this case, it's most readily pronounced from the transient (SM) DPA medium created during rehearsal (and we get a predicate adjective).

6.2 Ideas about predicative structures

blob

7 Extensions

7.1 Predicate noun phrases

7.2 Generics

8 Related computational models

Plebe *et al.* (2013) have a model that learns first nouns and then adjectives, following the trajectory in infants. Their model only learns adjectives after a working memory system is activated. The model is a hierarchically-structured collection of SOMs. One set of SOMs process visual information (shape and colour are processed in separate areas); another set process auditory information. The highest-level SOM represented PFC, where the visual and auditory streams are combined. There are two versions of the PFC SOM, one without recurrent connections, and one with recurrent connections (simulating the development of WM). The training data for the non-recurrent SOM consists of visually presented objects (from the COIL corpus) paired with a single noun, or a single adjective. The training data for the recurrent SOM is a visual object combined with an auditory noun, and then *after a delay*, an auditory adjective. (A bit odd, this. And definitely, no attempt to model IOR-type phenomena in the visual system.)

[7] During rehearsal, the *SM* DPA holds a unit that means 'dog' as a side-effect of RPC activation. I have to explain why this doesn't get pronounced. I think it's because (a) the word *dog* is more reliably associated with the *WM* DPA pattern, which is tonically active, and (b) there are *other* side-effects of initial RPC activation that relate to adjustment of the 'dog' template (size, shape etc) that are also tonically active. But this is still not watertight.

References

Abney, S. (1987). *The English Noun Phrase in its Sentential Aspect*. Ph.D. thesis, MIT, Cambridge, MA.

Aboh, E. (2000). DP licensing and snowballing movement. In *Proceedings of the NP-DP conference*, Antwerp.

Alexiadou, A., Haegeman, L., and Stavrou, M. (2007). *Noun Phrase in the Generative Perspective*. Mouton de Gruyter, Berlin.

Binder, J. and Desai, R. (2011). The neurobiology of semantic memory. *Trends in Cognitive Sciences*, **15**(11), 527–536.

Bolinger, D. (1967). Adjectives in english: attribution and predication. *Lingua*, **18**, 1–34.

Cabredo Hofherr, P. (2010). Adjectives: an introduction. In P. Cabredo Hoffher and O. Matushansky, editors, *Adjectives: Formal Analyses in Syntax and Semantics*, pages 1–28. John Benjamins, Amsterdam.

Cant, J. and Goodale, M. (2007). Attention to form or surface properties modulates different regions of human occipitotemporal cortex. *Cerebral Cortex*, **17**, 713–731.

Cant, J., Arnott, S., and Goodale, M. (2009). fMR-adaptation reveals separate processing regions for the perception of form and texture in the human ventral stream. *Experimental Brain Research*, **192**, 391–405.

Carnie, A. (1995). *Non-Verbal Predication and Head-Movement*. Ph.D. thesis, Department of Linguistics and Philosohy, MIT.

Cavina-Pratesi, C., Kentridge, R., Heywood, C., and Milner, A. (2010). Separate channels for processing form, texture, and color: Evidence from fmri adaptation and visual object agnosia. *Cerebral Cortex*, **20**, 2319–2332.

Chatterjee, A., Thomas, A., Smith, S., and Aguirre, G. (2009). The neural response to facial attractiveness. *Neuropsychology*, **23**(2), 135–143.

Chen, D., Lu, H., and Holyoak, K. (2014). The discovery and comparison of symbolic magnitudes. *Cognitive Psychology*, **71**, 27–54.

Cinque, G. (1994). Evidence for partial N-movement in the Romance DP. In G. Cinque, J. Y. Pollock, L. Rizzi, and R. Zanuttini, editors, *Towards Universal Grammar: Studies in Honor of Richard Kayne*. Georgetown University Press, Washington, DC.

Coutanche, M. and Thompson-Schill, S. (2015). Creating concepts from converging features in human cortex. *Cerebral Cortex*, **25**, 2584–2593.

Delsing, L.-O. (1993). *The internal structure of noun phrases in the Scandinavian languages*. Ph.D. thesis, Lund University.

Giusti, G. (1997). The categorical status of determiners. In L. Haegeman, editor, *The new comparative syntax*, pages 95–123. Longman, London.

Herbert, C., Ethofer, T., Anders, S., Junghofer, M., Wildgruber, D., Grodd, W., and Kissler, J. (2009). Amygdala activation during reading of emotional adjectives—an advantage for pleasant content. *Scan*, **4**, 35–49.

Heycock, C. and Kroch, A. (1997). Inversion and equation in copular sentences. In *Proceedings of the Workshop on (Pseudo)clefts*, Zentrum für Allgemeine Sprach-wissenschaf (ZAS), Berlin.

Ishizu, T. and Zeki, S. (2011). Toward a brain-based theory of beauty. *PLoS One*, **6**(7), e21852.

Kawabata, H. and Zeki, S. (2004). Neural correlates of beauty. *Journal of Neurophysiology*, **91**, 1699–1705.

Kemmerer, D. (2012). The cross-linguistic prevalence of sov and svo word orders reflects the sequential and hierarchical representation of action in brocas area. *Language and Linguistics Compass*, **6**(1), 50–66.

Komatsu, H. and Ideura, Y. (1993). Relationships between color, shape, and pattern selectivities of neurons in the inferior temporal cortex of the monkey. *Journal of Neurophysiology*, **70**, 677–694.

Larson, R. (1998). Events and modification in nominals. In D. Strolovitch and A. Lawson, editors, *Proceedings of SALT VIII*, pages 145–168, Ithaca, NY.

Longobardi, G. (1994). Reference and proper names: A theory of N-movement in syntax and logical form. *Linguistic Inquiry*, **25**(4), 609–665.

Macoir, J., Laforce, R., Brisson, M., and Wilson, M. (2015). Preservation of lexical-semantic knowledge of adjectives in the semantic variant of primary progressive aphasia: Implications for theoretical models of semantic memory. *Journal of Neurolinguistics*, **34**, 1–14.

Massam, D. and Smallwood, C. (1997). Essential features of predication in english and niuean. In *Proceedings of NELS 27*, pages 236–272.

McCarthy, R. and Warrington, E. (1988). Evidence for modality-specific meaning systems in the brain. *Nature*, **334**, 428–430.

Morita, T., Kochiyama, T., Okada, T., Yonekura, Y., Matsumura, M., and Sadato, N. (2004). The neural substrates of conscious color perception demonstrated using fMRI. *NeuroImage*, **21**, 1665–1673.

Moro, A. (1997). *The raising of Predicates: Predicative Noun Phrases and the Theory of Clause Structure*. Cambridge University Press, Cambridge, UK.

Murphey, D., Yoshor, D., and Beauchamp, M. (2008). Perception matches selectivity in the human anterior color center. *Current Biology*, **18**(3), 216–220.

Öhman, A. (2005). The role of the amygdala in human fear: Automatic detection of threat. *Psychoneuroendocrinology*, **30**, 953–958.

Olson, I. and Marshuetz, C. (2005). Facial attractiveness is appraised in a glance. *Emotion*, **5**(4), 498–502.

Patterson, K., Nestor, P., and Rogers, T. (2007). Where do you know what you know? the representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience*, **8**(12), 976–987.

Pessoa, L. (2005). To what extent are emotional visual stimuli processed without attention and awareness? *Current Opinion in Neurobiology*, **15**, 188–196.

Plebe, A., De la Cruz, V., and Mazzone, M. (2013). In learning nouns and adjectives remembering matters: A cortical model. In A. Villavicencio, T. Poibeau, A. Korhonen, and A. Alishahi, editors, *Cognitive Aspects of Computational Language Acquisition, Theory and Applications of Natural Language Processing,*. Springer, Heidelberg.

Pulvermüller, F. and Hauk, O. (2006). Category-specific conceptual processing of color and form in left fronto-temporal cortex. *Cerebral Cortex*, **16**, 1193–1201.

Wicker, B., Keysers, C., Plailly, J., Royet, J.-P., Gallese, V., and Rizzolatti, G. (2003). Both of us disgusted in *my* insula: The common neural basis of seeing and feeling disgust. *Neuron*, **40**, 655–664.