UNIVERSITY OF MURCIA

International Journal
of
English Studies

IJES

www.um.es/ijes

Sparseness and Entropy in Semantic Change: Precedents from Early Vision

HARRY HOWARD*
Tulane University

ABSTRACT

What is the difference between description and explanation in linguistics? We take explanation to be the reduction of description to independently motivated general principles. For language, a promising source of independently motivated general principles can be found in neuroscience, and especially computational neuroscience. As an example, we consider the notion of historical change in word meaning, which in cognitive semantics is often described in terms of conceptual metaphor, metonymy and synecdoche. Why does language rely on these three mental abilities and not some others? We argue that it is because the brain depends on the computational principles of sparseness and entropy to generate efficient mental representations. These principles have been found at work in primary and secondary visual cortex, so that a proper understanding of semantic change takes us through the physiology of early visual cognition.

KEYWORDS: semantic change, metaphor, metonymy, synecdoche, sparseness, entropy, visual cortex, computational neuroscience

---

****Address for correspondence***: Harry Howard. Tulane University. Dept. of Spanish & Portuguese. 322-D Newcomb may. New Orleans, LA 70118. USA. Tel: +504-862-3417; Fax: +504-862-8752; e-mail: howard@tulane.edu

IJES, vol. 7 (1), 2007, pp. 17-33

I. INTRODUCTION

Cognitive Linguistics has shown itself to describe certain linguistic data more accurately and parsimoniously than its alternatives. We believe that it has now reached a level of maturity in which it no longer needs to validate its utility by piling up more descriptive successes, but rather must turn to the explanatory challenge of showing *why* its mechanisms are correct, by deriving them from first principles. Just asserting their correctness on the strength of their descriptive adequacy is insufficient, for it leads to circularity. For instance, just asserting that such-and-such a semantic change relies on conceptual metaphor may be descriptively accurate, but begs the question of why semantic change relies on conceptual metaphor, as opposed to some other cognitive ability that humans have. Without an answer to this question, all we know is that language uses conceptual metaphor, because we see it in semantic change, and that conceptual metaphor is useful for language, because we see it in semantic change. In this paper, we reduce several properties of semantic change to independently-motivated principles of visual processing found in primary and secondary visual cortex, namely, the notions of the sparseness of visual representations and the development of a visual precept through decreasing entropy.

II. SEMANTIC CHANGE

One domain in which many people find Cognitive Linguistics to be particularly compelling is that of semantic change. By way of introduction, Brigitte Nerlich and David D. Clark assimilate semantic change to two general principles of semantic innovation. In Nerlich and Clark (1992:137), they put it so:

> the trick of being innovative and at the same time understandable is to use words in a novel way the meaning of which is self-evident … there are only two main ways of going about that: using words for the near neighbors of the things you mean (metonymy) or using words for the look-alikes (resemblars) of what you mean (metaphor).

In Nerlich and Clark (1999:128), they add a third 'leg', synecdoche, to form a triangle of semantic abilities, cf. Seto (1999):

> Metaphor is based on 'seeing similarities' (e.g. "She is the sun of my life"), metonymy is based on 'exploiting connections' (e.g. "I am giving a paper"), and synecdoche is based on understanding relations between categories, on 'understanding class inclusion' (e.g. "Give us our daily bread").

To put synecdoche in the more direct terms of the first quote, we could say that it is innovation using words for a different degree of specificity of the things you mean, though that does not lilt off the tongue as easily as Nerlich and Clark's other definitions do. Dirven and Verspoor (1998:42) summarize the three pragmatic abilities discussed by Nerlich and Clark as three kinds of conceptual relations, with the visual aid of Table 1, where we have inserted 'synecdoche' for the hierarchical relations that Dirven and Verspoor do not name on a par with metonymy and metaphor:

Conceptual relation	in semasiology	in onomasiology
Hierarchy (top /bottom)	generalizing and specializing [synecdochical extensions of senses]	conceptual domain: taxonomy and lexical field [conceptual synecdoche]
Contiguity (close to)	metonymical extensions of senses	conceptual metonymy
Similarity (like)	metaphorical extensions of senses	conceptual metaphor

Table 1. Conceptual relations in semasiological and onomasiological analysis, from Dirven and Verspoor (1998)

More recently, Traugott and Dasher (2002) attempt to reduce these three pragmatic abilities or conceptual relations to a single general format. They propose that inferences invited by a situation of use become generalized or conventionalized over time, thus moving the meaning of a grammatical form from one sort to the next. The way that inferences are invited by a situation of use involves the speaker/writer's ability to exploit metaphorization, metonymization (and synecdochization), but these abilities fall below the level of resolution of Traugott and Dasher's model.

III. SOME EXAMPLE OF SEMANTIC CHANGE

Dirven and Verspoor (1998:231-2) provides some simple examples of semantic change. In Middle English, the word 'bead', spelled 'bede' in Chaucer, had a more specific meaning than today, that of "one of the small perforated balls forming the rosary or paternoster, used for keeping count of the number of prayers said". Given that 'bede' is cognate with Dutch *bede*, *gebed* 'prayer' and German *Bitte*, *Gebet* 'request, prayer', Dirven and Verspoor propose the semantic change in (**¡Error! No se encuentra el origen de la referencia.**):

(1) bead: (i) prayer > (ii) ball standing for one prayer > (iii) string of balls forming the rosary > (iv) any string of balls used to wear as jewelry

These steps can explicated as follows. The step from (i) to (ii) is metonymy, from the frequent juxtaposition of saying a prayer and holding a bead. The step from (ii) to (iii) is synecdoche, generalizing from a single bead to all the beads on a string. The step from (iii) to (iv) is also synecdoche, generalizing from prayer beads to the beads found on any necklace.

Another example is in the distinction between 'dog' and 'hound' in Middle English. The basic-level term for canine was 'hound', as in contemporary Dutch *hond* and German *Hund* 'dog'. 'Dog' itself was reserved for a particular breed of canine, perhaps something like a mastiff. This breed was so prized that the name for it came to be the basic-level term for canine, while the reference of 'hound' was narrowed to that of the subtype or breed that we recognize today. We can schematize this change as in (**¡Error! No se encuentra el origen de la referencia.**):

(2) a. dog: *mastiff > … > domesticated canine
b. hound: domesticated canine > … > a dog of any of various breeds used in the chase and commonly hunting by scent

(**¡Error! No se encuentra el origen de la referencia.**) illustrates reciprocal synecdochization, as ‘dog’ is generalized to become the basic-level term, ‘hound’ is particularized to what ‘dog’ had been, a name for a breed.

A third example concerns the polysemy of the Middle English word ‘mete’. It was the basic-level word for any kind of food, as well as solid food as opposed to drink, or even a meal in general, especially dinner. It also referred to the edible part of fruits, nuts, eggs, and such, as well as the edible flesh of animals used for food, the only sense that is retained with any frequency today. We can summarize these meanings as in (**¡Error! No se encuentra el origen de la referencia.**), which tries to maintain the visual format of the previous examples, even though a radial category would be more perspicuous

(3) a. mete: food > Ø
b. mete: solid food, not drink > Ø
c. mete: meal in general, especially dinner > Ø
d. mete: edible part of fruits, nuts, eggs, and such > Ø
e. mete: the flesh of animals used for food > … (meat)

If (**¡Error! No se encuentra el origen de la referencia.**a) is the source usage, then (**¡Error! No se encuentra el origen de la referencia.**b) would be derived by metonymy (food being frequently associated with drink in a meal), (**¡Error! No se encuentra el origen de la referencia.**c) would be derived by synecdoche (food being part of a meal) and (**¡Error! No se encuentra el origen de la referencia.**d) and (**¡Error! No se encuentra el origen de la referencia.**e) would also be derived by similar sorts of synecdochization for the edible part of a foodstuff and for the edible part of an animal, respectively.

Traugott and Dasher (2002) review semantic change in several syntactic categories. One that is fairly simple to illustrate is that of the development of three adverbials, plus one modal verb, in English. Traugott and Dasher state some of their results in terms of the cross-linguistic hierarchy of adverb senses proposed in Cinque (1999). We have organized this hierarchy into Table 2 so as to plot the development of the four lexical items taken from Traugott and Dasher as ascension in Cinque’s hierarchy. The years indicated under the items refer to the approximate dates of initial usage in the corpus:

Modal sense	**Adverb example**	***must***	***indeed***	***actually***	***in fact***
Hedge	well	-	-	-	-
Discourse marker	then	-	1600	1815	1815
Speech act	frankly	-	↑	↑	↑
Evaluative	fortunately	-			
Evidential	alleged	-			
Epistemic	probably	1300	1450	1750	1680
Past	once	↑	↑	↑	↑
Future	then				
Irrealis	perhaps				
Necessity	necessarily				
Possibility	possibly				
Volitional	intentionally				
Obligation		1000			
Ability/Permission		700’s			
[Manner]		-	1300	1425	1670

Table 2. Change in English adverb and modal verb meanings, from Traugott and Dasher (2002)

As an illustrative example, consider ‘indeed’. Traugott and Dasher find that in Old English, ‘dede’ is a full noun meaning ‘action’ which can be modified by demonstratives, adjectives, and quantifiers. It is also used as a bare noun with ‘in’ as an adverbial of manner, meaning ‘in action’, often in contrast to ‘in speech’ or ‘in thought’. Traugott and Dasher argue that if something happens ‘in action’, it invites the inference that the event is observable, through what they call the “fallacy ‘seeing is believing”, (p. 160), a “fallacy” that would be called a conceptual metaphor in cognitive semantics. Through this inference, ‘in dede’ comes to bear evidential or epistemic weight: ‘in action/practice’ > ‘in actuality, certainly’. And indeed, around 1450 the first usages of ‘in dede’ are found expressing the speaker/writer’s commitment to the truth of the proposition. If epistemic ‘in dede’ indicates commitment to the veridicality of its proposition, the inference can be drawn that the statement of its proposition is of a better or more appropriate form than some other statement, usually the preceding one in the discourse. This inference is perhaps a conceptual metaphor – asserting the truth of p is asserting the well-formedness of p – but we know of no such conceptual metaphor, and trying to justify it would take more thought than we can dedicate to it here. The 1600’s bring the first attestations of the conventionalization of this inference for ‘indeed’. We snuck an example of it into our text four sentences back, and the reader can appreciate how the present-day ‘indeed’ acts as a clause-initial discourse marker.

IV. STEPS TOWARDS A THEORY OF SEMANTIC CHANGE

Shorn of all their particularities, the theories and examples of semantic change reviewed above can be schematized as in Fig.1

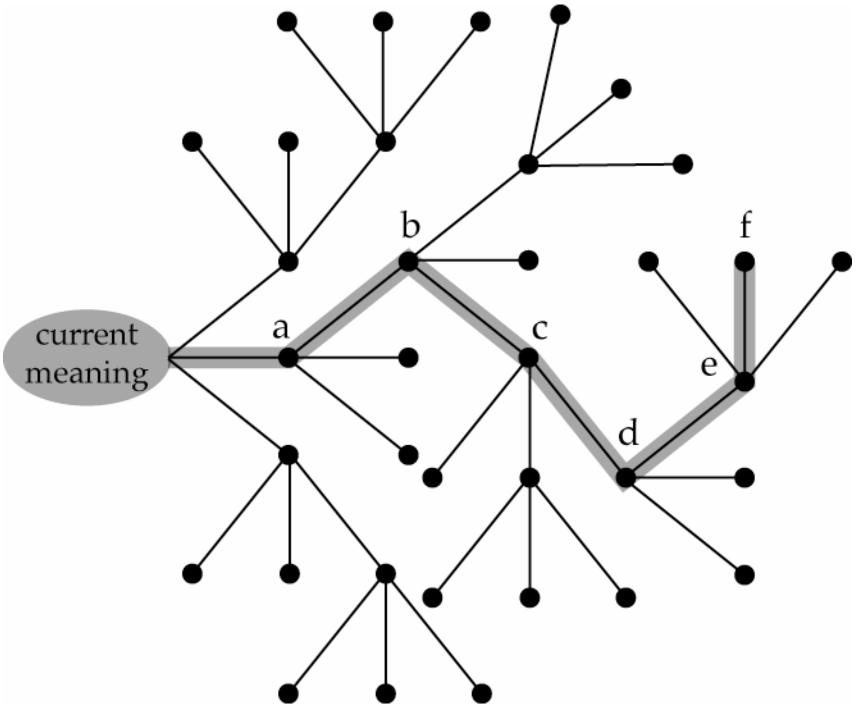

Figure 1. Hypothetical pathway of semantic change through a network.

In this figure, each line (technically, an edge) represents one of the three pragmatic abilities or invited inferences, and each filled-in circle at the end of such a line (technically, a node) represents an inferred meaning. The wide gray line traces a particular hypothetical path of change. For instance, the path from *a* to *b* could be established by metaphorization, that is, a similarity in quality of *b* with *a*; the path from *b* to *c* could be established by metonymization, that is, by *c* being contiguous with *b*; and the path from *c* to *d* could be established by synecdochization, that is, *d* can be found by following the relations of subordination or domination in a hierarchy that includes *c*.

We do not mean to advance any great claim to originality with this proposal. The literature on metaphor, metonymy, and synecdoche is voluminous; someone may have already advanced a reduction of semantic change to a topological structure such as Fig.1, and certainly the format of Fig. 1 is compatible with, if not implicit in, a variety of cognitive-semantic representations. Where we can make a claim to originality, however, is to propose a neurological underpinning for the reduction seen in Fig. 1.

There are two questions that we wish to answer about this figure. The first is whether there is any relationship common to all of its edges, or in non-topological terms, to metaphor, metonymy, and synecdoche. The second is why semantic change follows but a single path, instead of following them all simultaneously.

IV.1. Semantic change and entropy

To answer the first question, we think that if there is a relationship common to all the edges of Fig. 1 – it is **entropy**, particularly Shannon's entropy, which is a measure of the predictability or surprise value of a response, see Shannon (1948) and voluminous posterior work. Frequent responses come to be expected and so are more predictable or less surprising.

Infrequent responses are not expected and so are less predictable or more surprising. This notion is usually illustrated by the amount of information gained from tossing a coin or receiving a symbol, but such stories still seem lifeless and opaque to many people, so we will make up an illustration from the data just seen.

Say we are speakers of Middle English, and you tell me you have one more 'bede' to go, while holding your rosary in your hand. What do I think that you mean? There is always some uncertainty about what words mean, and in this particular case (as well as all of the others that we have discussed) there are three possibilities. The first is that by 'bede' you meant 'prayer'. Since this is what we have taken to be the meaning of 'bede', it is likely that it is what you meant. Let us evaluate this interpretation as the highest measure of probability, which by convention is called one, or $p = 1$. But maybe you meant 'grandmother'. Well, there is no obvious way to understand 'grandmother' for 'prayer', so that seems improbable. Let us evaluate this interpretation as the lowest measure of probability, which by convention is called zero, or $p = 0$. But, then again, I know you are keeping track of your prayers by counting the beads on your rosary, so much so that the two are practically equivalent acts, so I may infer that you may mean 'bead', a metonymy. In fact, I may infer this even if what you meant was 'prayer' and not 'bead'. So understanding 'bead' for 'bede' has some probability intermediate between 1 and 0. Let us call it 0.5 for the sake of convenience.

It may seem that we are belaboring the obvious, but if understanding a word is a question of probabilities, what theory of semantic change can we build on it? If we say that we always choose the most probable meaning ($p=1$), then we correctly reject the improbable ones ($p=0$), but we also incorrectly reject the middling probable ones ($p=0.5$), the pragmatic inferences. We wind up with no change. Words are always understood to have their most probable meanings. We therefore conclude that probability supplies the wrong theory of semantic change.

There is an alternative. We want to privilege the middling probable ($p=0.5$) over the highly probable ($p=1$), but we still want to reject the improbable ($p=0$). Entropy is the way. It says that the highly probable is too certain to happen to be interesting, while the highly improbable is too certain to not happen to be interesting. What is interesting, or surprising, is the middling probable, the zone of uncertainty, and that is the measure that is elevated over the other two. Mathematically, this is done by taking the negative of the log of the probability, which is graphed for equal probabilities as in Fig.2.

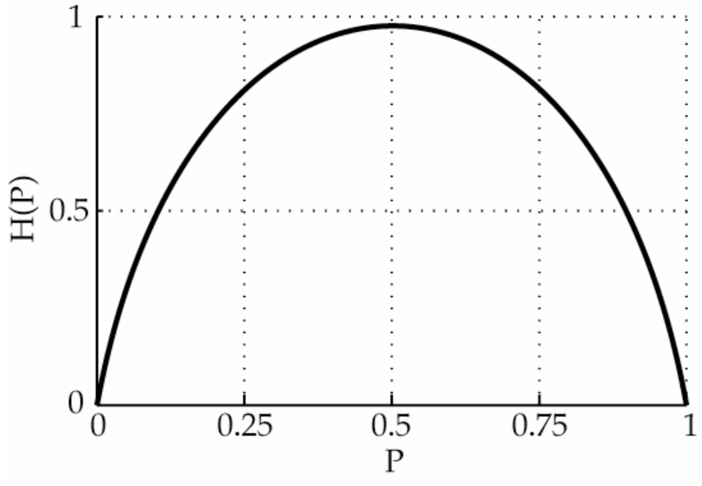

Figure 2. Entropy, H, as a function of probability, P

With this measure, we can define the direction of semantic change as that which increases the entropy of the word. Conserving the same meaning gives an entropy of zero, H = 0 in Fig.2, and, of course, the word does not change. Jumping to a random meaning also gives an entropy of 0, and, as far as we know, words do not change randomly. Rather, word meanings develop in a fashion motivated by the pragmatic triangle, which produces an entropy near 1.

Note how efficient this is. The semantic change system does not have to devote resources to tracking meanings that do not change, nor does it have to devote resources to anticipating random, unmotivated changes. It focuses its resources on precisely those potential changes which have the greatest chance of propagating through a community.

There is one nagging doubt that we have about the proposal, which should be aired before moving on. It is that we have relied on probability to calculate entropy. We did this because it follows Shannon's practice, which has proved to be too useful to change. Yet Shannon was studying the most abstract case, that of a message sent along a channel between any emitter and any receiver. In such a state of ignorance about the make-up of the emitter and the receiver, the best one can do is to make probabilistic statements. Linguists, on the other hand, know the emitters and receivers that interest them very well. That is, we do not have to content ourselves with statements as vague as "the probability that 'bede' means 'prayer'" because we do not know what causes 'bede' to mean 'prayer'. We do know what causes 'bede' to mean 'prayer': it is the correlation, or association, between the phonological form 'bede' and the concept 'prayer' acquired by native speakers of Middle English. Let us agree to be precise when we can be, and substitute the notion of correlation for that of probability. It goes without saying that correlation must be measured on the same scale: highly correlated, i.e. identical, meanings must evaluate to 1, partially correlated meanings must evaluate to 0.5, and uncorrelated meanings must evaluate to 0.

With this notion of correlation, we can restate the pragmatic triangle in a most accurate way. Metaphor links the source and the target domains by means of a correlation in similarity; metonymy links the source and the target domains by means of a correlation in space and time, and synecdoche links the source and the target domains by means of the correlations established among the levels of a hierarchy. Inferences are invited by finding such correlations. It is up to the pragmatic abilities of the speaker/writer to find them, and then it is up to the hearer/reader to recognize them and repeat them enough so that they become entrenched in the community.

IV.2. Semantic change and sparseness

The second question that we want to ask about Fig. 1 is why semantic change only follows a single path from one node to the next. The reader may have an immediate, knee-jerk answer, namely, how could it be otherwise? Well, it could be otherwise. Under Shannon's definition of a communication channel, a channel is used most efficiently when all of its resources are used at the same time, see for instance Graham and Field (2006). If we take the transition from one node to the next to be a Shannon channel, then its most efficient utilization implies that all of the edges that exit it be followed and used at the same time. Yet semantic change does not fan out indiscriminately from a node; instead, it strikes out in single file, along a single edge.

The reader may surmise that there is some limitation on semantic or attentional resources that impedes a Shannon-style fan out, and this may ultimately turn out to be true. However, there is a fact that we have glossed over, namely that many instances of semantic change follow a similar sort of development, from specific to general. That is the case of all of the changes summarized in Table 2 from Traugott and Dasher (2002). This sort of change is so common that Traugott (1997:185) labeled it **subjectivization**, defining it as "the historical pragmatic-semantic process whereby meanings become increasingly based in the speaker's subjective belief state, or attitude toward what is said".

Our answer to the second question is to propose that semantic change is **sparse**. By sparse, we refer to a representational scheme in which only a few units (out of a large population) are used at any time. If the reader recalls Fig. 1, it has 34 units (nodes), each representing a potential new meaning reachable from the current one. However, the trajectory of semantic change only uses six of those units, and for any given mention of the item, only one node would be used, out of 34. Thus semantic change is sparse, making it very different from, and much less efficient than, a Shannon channel. Sparseness has other advantages, however, one of which is that it automatically accounts for the subjectivisation effect of semantic change, though to explain this proposal adequately, we must take a long detour into the brain.

V. EARLY AND INTERMEDIATE VISION

It is convenient to separate the mammalian visual system into two halves, the first starting at the retina, passing through the thalamus at the lateral geniculate nucleus (LGN) ending in primary visual cortex (V1), and the second including all visual processing after V1. This bifurcation is depicted in Fig.3.

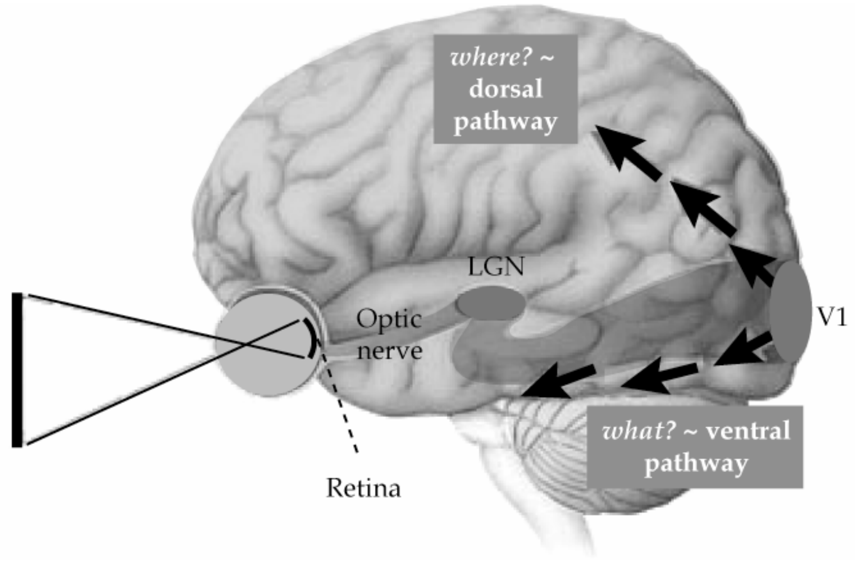

Figure 3. Streams of visual processing

In the first half, or early vision, visual information is processed in the same way. In the second half, late vision, visual information splits into that which is necessary to recognize an object, the *what?* or ventral pathway, and that which is necessary to locate an object in space, the *where?* or dorsal pathway.

V.1. V1 and sparseness

Recordings reported in Hubel and Wiesel (1962) of the electrical response of single neurons in V1 revealed them to be sensitive to lines and edges. In particular, a simple V1 cell responds strongest to a bar of light (or darkness) that has a particular orientation and position in visual space. This is a striking discovery, since the input from the LGN does not have any such selectivity.

From many such experiments, Hubel and Wiesel concluded that the cells that they were studying must gather outputs from a lozenge-shaped group of LGN cells. This group of input cells is called the receptive field of the V1 cell. The receptive field that creates the responses illustrated in Fig. 4 is illustrated in Fig. 5 on the next page.

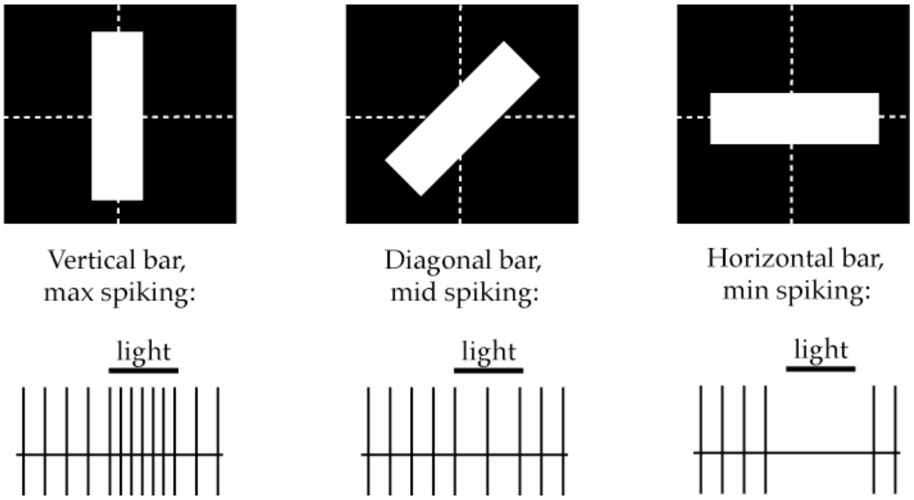

Figure 4. Response of a single simple V1 neuron to a bar of light according to its orientation

Note that its center bit is sensitive to light while the surrounding dark edges are sensitive to darkness.

Despite the fundamental nature of this discovery, an explanation for it (i.e. reduction to general principles) was not forthcoming for another thirty years, when statistical analysis of a sample of images of the natural world reported in Olshausen and Field (1996) showed that their simplest components are oriented lines. What Olshausen and Field did was to divide 512-by-512 pixel photographs of natural objects into smaller, 12-by-12 pixel patches and then randomly create simpler patches of the same size, called **basis functions**, that could be added together to construct a real patch. Olshausen and Field designed an algorithm that adjusts the basis functions incrementally as many thousands of patches are presented to them, so that on average the smallest possible number of them can reconstruct each entire image. As Olshausen and Field (2000:244) put it, "… the algorithm seeks a "vocabulary" of basis functions such that only a small number of "words" are typically needed to describe a given image, even though the set from which these words are drawn might be much larger." A representation in which only a few such 'words' are used at any given time is sparse.

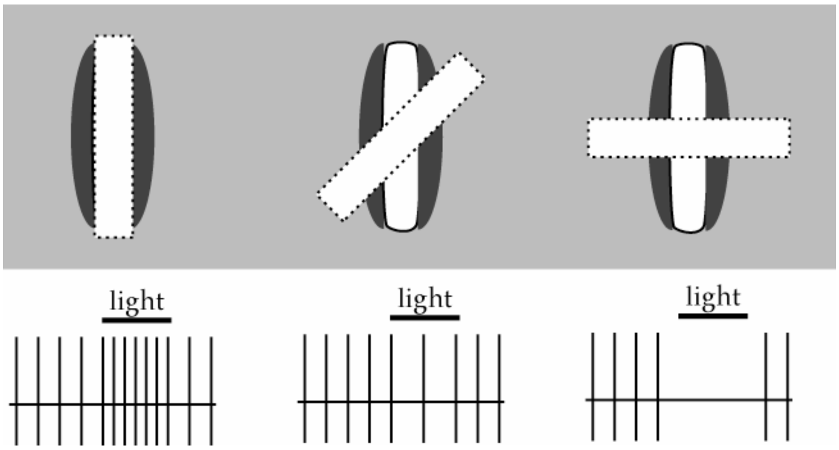

Figure 5. A receptive field proposed by Hubert and Wiesel for the simple V1 cell response

The basis functions that emerged from this process are reproduced in Fig. 6 [Next page]. Note how they match the lozenge shape seen in the previous figure, though at different orientations, and some of them have the center bit being sensitive to darkness and the surrounding bits to light, which is the opposite of Fig. 5. These functions are usually called oriented line segments, but given the fact that recent work indicates that they can organize in V1 into groups that form the contour lines of objects (see Field and Hayes (2004) for review), we refer to them as contour segments. The fact that such contour segments emerged without Olshausen and Field imposing any other constraints or assumptions, suggests that neurons in V1 are indeed configured to represent natural scenes in terms of a sparse code.

Why would this be a good thing to do? The best way to understand the advantages of sparse representation is to consider the difference in how the LGN and V1 represent a given stimulus, such as the dark bar in Fig.7. On the left side of the figure, are depicted the overlapping bull's-eye-shaped LGN receptive fields along the bottom, which project up to the dots representing the neurons at the top. Note that the fact that the dots are filled-in means that all three neurons are active. On the right side of the figure, a more complex receptive field neuron projects up to a single V1 neuron; its two neighbors do not receive the proper stimulus and so are inactive.

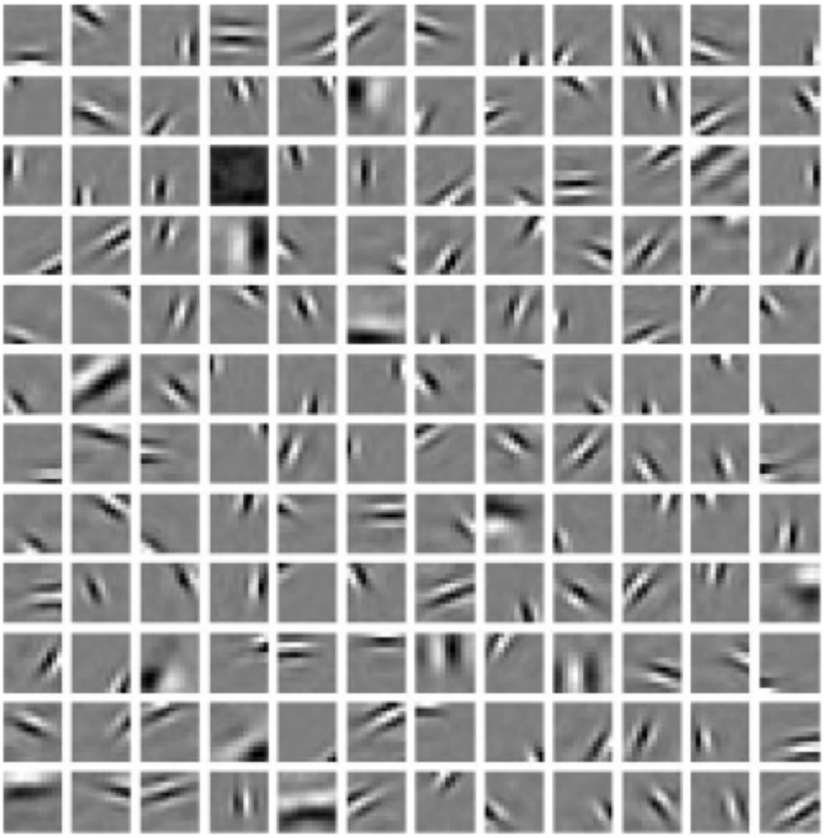

Figure 6. Basis functions of a large sample of natural images

Both sample neural populations receive the same input, but V1 represents it much more concisely, by the output of a single neuron. This concision can simplify the job of higher levels of processing. As Olshausen (2003) says, “one way of potentially achieving a meaningful representation of sensory information is by finding a way to group things together so that the world can be described in terms of a small number of events at any given moment. In terms of a neural representation, this would mean that activity is distributed among a small fraction of neurons in the population at any moment, forming a sparse code.” And that is just what we saw in V1: a neuron will respond to a contour segment at a particular orientation. The neurons that are sensitive to other orientations in the same patch will be silent.

Returning to this paper’s topic of semantic change, we can suppose that it hews to the purpose of sparseness of helping out higher levels of processing by describing the world in terms of a small number events. We can draw the further conclusion that the sparse representational format that we attribute to it is a general property of neurological architecture and not a special property of language. Thus we can simplify our model of the mind by not having to stipulate sparseness separately for semantic change. This is in line with the cognitive-linguistic program of reducing facts about language to general principles of cognition.

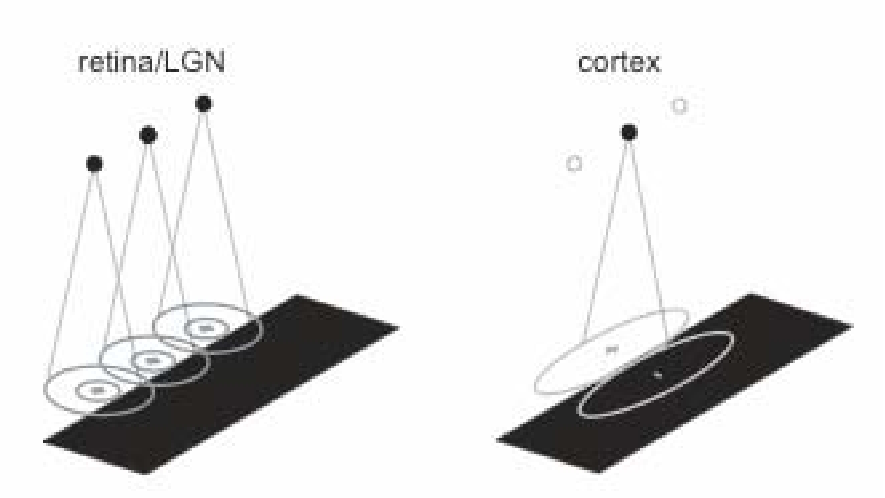

Figure 7. LGN vs. simple V1 tiling of a dark bar, after Olshausen (2003), Fig. 1

More interesting is the way in which the picture of sparsification in Fig. 7. shows the development of general representation from a more specific one. The LGN format on the left is more specific in that it needs three units to encode the stimulus. The simple V1 encodes the same stimulus with only a single unit. This is practically the definition of subjectification, so that we can conclude that the tendency towards generalization seen in semantic change is an expected by-product of a sparse encoding.

V.2. V2 and entropy

The next step ‘up’ in visual processing from V1 is nick-named, conveniently, V2, though it is more accurately referred to as Brodmann Area 18, the central part of the occipital cortex. V2 shares some of the properties of V1, such as orientation and spatial frequency, but new properties begin to emerge. The one that is most interesting for us is the fact that it is at V2 that one observes the resolution of visual objects into figure and ground. By the far the most well-known illustration of this ability is seen in Fig.8, where there are two possible resolutions. Known as the Rubin vase, it is one of several optical illusions developed by the Danish psychologist Edgar Rubin and published in 1915. It is commonly seen as two different images, a white vase on a dark background, or two dark faces facing each other against a white background.

Figure 8. Bistable image. What do you see?

The phenomenon underlying this kind of illusion is known as border ownership. The gist of it is whether the border between the light and dark parts of the images belongs to the light image, thereby forming a vase against a dark background, or to the dark images, thereby forming two faces against a light background. Zhou et al. (2000) observe neurons in V2 which are sensitive to border ownership, and Qiu and von der Heydt (2005) flesh out the hypothesis that V2 adds to the contour segments that enter from V1 an indication of border ownership, e.g. a given segment belongs to whatever is on side or the other of it. Zhaoping (2005) develops a computational simulation in which segments that would go together to make a closed figure facilitate one another, while other segments inhibit one another. The group dynamics that emerges from this architecture tends to create closed figures, if they are possible, as Zhaoping's simulations show.

For a bistable image such as Rubin's vase, Zhaoping's model tends to choose one interpretation over the other. This may reflect limitations of the model, or it may reflect the influence of other, currently unknown, factors. However, what is interesting to us is that the computational simulation shows how the neurons in V2 construct a percept by moving from a disordered state in which ownership of the various segments is random or at least not coherent, to an ordered state in which as many segments as possible are recruited by ownership to the same figure. Does this sound like semantic change?

It does from the perspective of entropy. Let us assume that the combination of LGN input into contour segments in V1 produces a state of order in V1, which is a state of low uncertainty or low entropy. However, when these segments are tagged for random border ownership upon entry to V2, this new state is highly uncertain, or high in entropy. The self-organizing process of facilitation and inhibition in V2 reduces this disorder to the most coherent interpretation possible given the input, which is a state of relative certainty or low entropy. This process is analogous to that of semantic change, that is, the development of meaning from a low entropy state via a high entropy pragmatic inference to an entrenched low energy state of new meaning.

This reasoning can be extended to account for the perception of bistable images such as Fig. 8. One's initial interpretation, say as mirrored faces, is a low-entropy state produced in V2. For whatever reason, perhaps attention to a certain part of the image, V2 is forced to reorganize its allocation of border ownership, which produces an intermediate state of high entropy as uncertainty increases and a new interpretation is searched for. The interpretation as a vase emerges, allowing entropy to fall back down to a low state. The three processes are compared in Table 3.

Entropy	**Semantic change**	**Figure/ground segmentation**	**Perception of bistable image**
low entropy	current meaning	analysis into contour segments in V1	initial interpretation
high entropy	motivated change	contour segments enter V2 with random ownership	switch (no interpretation)
low entropy	entrenched new meaning	self-organizing border ownership in V2	other interpretation

Table 3. Comparison of semantic change and intermediate vision in terms of entropy

Thus we once again do not need to stipulate a sensitivity to entropy in semantic change, since it appears to be a general property of cerebral functioning.

VI. CONCLUSIONS

In this paper, we have tried to reduce two general properties of semantic change to two general properties of neocortical functioning, as found in early visual processing.

We have argued that semantic change is sparse and therefore very different from the notion of an efficient communication channel studied in information theory. By being sparse, semantic change will tend go from specific meanings to general meanings, which is exactly what has been documented for the kind of semantic change known as subjectivisation. We were inspired to look for sparseness in semantic change from the well known postulation of sparseness in the organization of contour segments in V1.

We have also argued that semantic change proceeds from a state of low entropy (the current meaning of the item), to a state of high entropy (a pragmatic inference to a new meaning), then back to a state of low entropy (once the new meaning becomes entrenched in a community). We have found a precedent for this appeal to entropy in figure-ground segmentation in V2, in which contour segments output by V1 self-organize into figures in terms of border ownership.

The reader may still be skeptical that principles of low-level perceptual processing can have any bearing on a high-level cognitive phenomenon like semantic change. However, there is actually a very good reason for this to be the case. It is a conjecture that has inspired our research for more than the last ten years, which has come to be called the **isocortex conjecture**, stated in (1):

1. Given that all perceptual, cognitive, and motor faculties are subserved by an architectonically similar six-layer neocortex (see Rockel, Hiorns and Powell (1980)),
   a) the computations that subserve cognitive faculties should be similar to those performed in visual cortex, just over different primitives;
   b) the computations performed at every level in a cortical hierarchy should be similar, allowing for the difference in primitives.

We choose the term ‘isocortex’ to refer to cerebral cortex in this conjecture, because it underscores the architectonic similarity of all areas of the cerebrum, as opposed to the more common term ‘neocortex’, which underscores the evolutionary recency of the cerebrum, which does not interest us.

The isocortex conjecture licenses us to look for explanation of cognitive phenomena, whose cortical substrate is often unknown, in motor and perceptual processing, whose cortical substrate is often much better known, since it can be studied (more or less) ethically in non-human animals. Moreover, from a meta-theoretical perspective, the isocortex conjecture enforces a notion of parsimony on the global theorizing of the human sciences, namely, that it does not help to invent a theoretical notion for linguistics that has no analog in any other aspect of the perceptuo-cognitive-motor system, since doing so may make linguistic theory simpler, but it makes the mind more complicated. Note that this meta-theoretical parsimony may actually be grounded in an actual evolutionary parsimony, in that the linguistic regions of the brain presumably developed from cortical regions that were already doing something else, so that all language did was add an new function to existing machinery.

We close by expressing our hope that linguistics is nearing the day when it has the rigor of a natural science, in which observations are reduced to the interaction of independently-motivated primitives.

REFERENCES

Cinque, G. (1999). *Adverbs and Functional Heads: A cross-linguistic perspective*. Oxford: Oxford University Press.

Dirven, R. & Verspoor, M. (1999). *Cognitive Exploration of Language and Linguistics*. Philadelphia: John Benjamins.

Field, D. J. & Hayes, A. (2004). Contour integration and the lateral connections of V1 neurons. In L. M. Chalupa & J. S. Werner (Eds.), *The Visual Neurosciences*.. Cambridge, USA: The MIT Press, pp. 507-521.

Graham, D. J. & Field, D. J. (2006). Sparse Coding in the Neocortex. In J. H. Kaas & L. A. Krubitzer (Eds.), *Evolution of Nervous Systems*. Amsterdam: Elsevier, pp. 181-188.

Hubel, D. H. & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology*, 160, 106-154.

Nerlich, B. C., David D. (1992). Outline of a model for semantic change. In G. Kellermann & M. Morrissey (Eds.), *Diachrony without Synchrony: Language history and cognition (Papers from the International Symposium at the University of Duisburg, 26-28 March 1990)*. Frankfurt: Peter Lang, pp. 125-144.

Nerlich, B. C., David D. (1999). Elements for an Integral Theory of Semantic Change and Semantic Development. In R. Eckardt & K. v. Heusinger (Eds.), *Meaning Change – Meaning Variation. Workshop held at Konstanz, Feb. 1999*, Vol. I. Arbeitspapier 106. Konstanz: Fachbereich Sprachwissenschaft, Universität Konstanz, pp. 123-134.

Olshausen, B. A. (2003). Principles of image representation in visual cortex. In L. M. Chalupa & J. S. Werner (Eds.), *The Visual Neurosciences*. Cambridge, USA: The MIT Press, pp. 1603-15.

Olshausen, B. A. & Field, D. J. (1996). Emergence of simple cell receptive field properties by learning a sparse code for natural images. *Nature*, 381, 607-609.

Olshausen, B. A. & Field, D. J. (2000). Vision and the coding of natural images. *American Scientist*, 88, 238-224.

Qiu, F. & von der Heydt, R. (2005). Figure and ground in the visual cortex: V2 combines stereoscopic cues with gestalt rules. *Neuron*, 47(1), 155-166.

Rockel, A.J., Hiorns, R.W. & Powell, T.P. (1980). The basic uniformity in structure of the neocortex. *Brain*, 103(2), 221-244.

Rubin, E. (1915). *Synsoplevede Figurer. Studier i psykologisk Analyse*. Copenhagen: Glydenulska.

Seto, K.-i. (1999). On distinguishing metonymy from synecdoche. In K.-U. Panther & G. Radden (Eds.), *Metonymy in Language and Cognition*. Amsterdam/Philadelphia: John Benjamins, pp. 91-120.

Shannon, C. E. (1948). The mathematical theory of communication. *Bell System Technical Journal*, 27, 379-423 and 623-656.

Traugott, E. C. (1997). *Unless* and *but* conditionals: a historical perspective. In A. Athanasiadou & R. Dirven (Eds.), *On Conditionals Again*. Amsterdam & Philadelphia: John Benjamins, pp. 145-167.

Traugott, E. C. & Dasher, R. B. (2002). *Regularity in Semantic Change*. Cambridge, U.K. & New York: Cambridge University Press.

Zhaoping, L. (2005). Border ownership from intracortical interactions in visual area V2. *Neuron*, 47(1), 143-153.

Zhou, H., Friedman, H. S., & von der Heydt, R. (2000). Coding of border ownership in monkey visual cortex. *Journal of Neuroscience*, 20(17), 6594-6611.