DRAFT (UNDER DEVELOPMENT: COMMENTS WELCOME)
Spatial prepositions as higher order functions:
And implications of Grice's theory for evolution of language.
(A discussion note)
Aaron Sloman

First created: Sun Oct 8 2006
Last updated: 29 Sep 2009; 13 Aug 2010; 7 Oct 2018 (re-formatted) ;27 Feb 2019 (minor tidying)

Two related slide presentations are available here:
What evolved first: Languages for communicating, or languages for thinking (Generalised Languages: GLs)? (PDF)
http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#glang
Presented to Language and Cognition Seminar, School of Psychology, University of Birmingham.
19th Oct 2007

What is human language? How might it have evolved? (PDF)
http://www.cs.bham.ac.uk/research/projects/cosy/papers/#pr0702
(Birmingham AINC Presentation 5 Mar 2007)

This document is available in two formats:
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/spatial-prepositions.html
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/spatial-prepositions.pdf

---

NOTE: Parametric polymorphism

Since writing this, around 2008, I have realised that the concept of parametric polymorphism, as used in connection with 'Object Oriented' programming languages and higher order functional programming languages, is relevant to the remarks below about context sensitivity and the need to extend notions of compositional semantics. When I get time I shall rewrite this paper to take account of this.

See also the role of polymorphism in the discussion of consciousness here:
http://www.cs.bham.ac.uk/research/projects/cogaff/09.html#pach
Phenomenal and Access Consciousness and the "Hard" Problem: A View from the Designer Stance
International Journal of Machine Consciousness, 2010.

CONTENTS (Provisional)

Abstract (Revised 12 Mar 2007)

This discussion note suggests that some forms of expression that are apparently vague, inviting interpretations of their meaning in terms of probability distributions, would be better construed as having a different form of semantics, namely specifying an 'higher order' function from contexts to truth-conditions, where a 'context' has two components, namely the current situation and a current set of goals of speaker and hearer. So statements made using them have a two level semantics. The first level specifies the function, which has to be applied to arguments extracted from the context, which may be linguistic or non linguistic, including the purpose of the communication. Then when that function is applied to the arguments the result is a specification of truth-conditions (another function, from situations to truth-values). This can be extended to how questions and imperatives using those expressions also need to be interpreted.

That is a theory of the semantics of some vague expressions in *adult* language. During very early child language learning the semantics of those expressions is probably less sophisticated, and more closely tied to examples of the application of the expressions. It takes time to learn the appropriate generalisations. This paper says nothing about how that learning proceeds. However that progression would be part of a general process of developing higher order abstractions --- as outlined in:
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/meta-configured-genome.html
The Meta-Configured Genome

This is a special case of a much more general feature of the semantics of natural language: the meaning of a complex expression is typically a function not only of the structure and components of the expression (linguistic inputs) but also of aspects of the environment(s) of speaker and hearer and their communicative intentions or purposes, and related beliefs (which may or may not be shared).

I proposed this sort of interpretation for statements using 'better' in 1969 in
How to derive "Better" from "is", *American Phil. Quarterly* Vol 6, pp43--52,
http://www.cs.bham.ac.uk/research/cogaff/sloman.better.html
and soon after extended it to 'ought' and 'should'.
http://www.cs.bham.ac.uk/research/cogaff/sloman.better.html

But I think the phenomenon is much more common than has been realised. I try to show how the use of such semantic functions taking linguistic and non-linguistics arguments can be predicted on the basis of Grice's theory of communication, and draw some conclusions regarding the evolution of language, and the relations between linguistic and non-linguistic mental functions.

From this viewpoint, communication (verbal or non-verbal) is creative, mostly, but not always, collaborative, problem-solving, not the transmission and decoding of some signal or the transmission of mental content from one individual another. So the ability to use a language is just a special case of a more general ability to solve problems by combining different kinds of competence.

This is related to the amazing invention of a sign language by Nicaraguan deaf children and to arguments for the evolution of internal richly structured languages prior to the evolution of language for communication. This is a discussion paper and everything is still tentative. Comments and criticisms welcome.

> A talk based on a subset of these ideas was given at the University of Birmingham on 5th March 2007, available in PDF format here:
> What is human language? How might it have evolved?
> http://www.cs.bham.ac.uk/research/projects/cosy/papers/#pr0702

Introduction: vague indexicals

It is well known that words like, 'this', 'that', 'here', 'there' can be used in a context where what is being indicated is not defined solely by the words used, so that some aspect of the context of utterance has to be used to interpret what is said. This context could be linguistic, e.g. what was said in previous utterances, or non-linguistic, e.g. the direction of a pointing gesture, the direction of gaze or a nod, or, more subtly, the implicitly understood purpose of the communication. Also the purpose of communication is relevant to determining what 'here' refers to. E.g. if I am telling someone where to park a car and point, saying 'here', what is communicated will be different from the effect of pointing and saying 'here' if answering a question like 'Where did you last see your dog?'

I don't know if there is anyone who, as a result of a failure to understand this point about the context sensitivity of the meaning of 'here', has concluded that 'here' refers to a set of possible

regions of space with a probability distribution, whose parameters are discovered by doing doing psychological experiments, e.g. a psychological study to find out which region of a table top falls within the intended denotation of 'here' uttered by someone pointing at an empty table and saying, something like 'I left it here', or 'Please put it here', or 'Was it here?'. I expect an experimental setup could be used to induce a probability distribution over the surface of the table, with the highest probability at the point of intersection of the direction of the extension of the pointing gesture and the surface of the table, and the probability falling off radially in all directions (with a circular or non-circular distribution).

If you ask people questions, even totally pointless questions, you can get answers (especially in experimental conditions, e.g. students being paid per hour), which then provide data that can be fed into statistics packages. It is not clear that anything can be inferred from this about what is going on when the same form of words is used in a *real* communicative context. I expect this point is fairly obvious in relation to the words 'here' and 'there'. However, it is not so obvious in other contexts used for psychological experiments.

Spatial prepositions

Spatial prepositions and prepositional phrases like, 'above', 'below', 'to the left of', 'to the right of', 'in front of', 'behind', 'this side of', 'on the far side of', have a similar lack of determinacy out of context. There are two levels of indeterminacy, the first of which concerns specifying *whose* or *what's* left, front, back, etc. is in question, and there are many different cases to be discussed, depending on the sort of entity the preposition is applied to.

For example, whereas 'behind Fred', 'behind the car', 'behind the chicken' all admit the possibility of a front-back axis determined by Fred, the car or the chicken, 'behind the ball', and 'behind the wall', do not, though perhaps 'behind the house' does. If the object has no intrinsic front/back or left/right opposition, then the relevant subdivision of space can be made relative to the speaker, or a hearer, or a third person referred to in the utterance.

But even after the frame of reference has been determined, there can still be large areas, or large volumes, that count as being in front of or behind, or to the left of or to the right of, or above or below, something. The same indeterminacy applies to 'between' when used to specify a spatial region rather than an interval in a linear ordering, for instance in the phrase, 'between the house and the gate', where this could be interpreted as referring to the region in the rectangular projection from the gate to the house, or the parallelogram from the gate to the front door, or the quadrilateral determined by the gate and the front wall of the house, or even the region of a possibly curving path from the gate to the house. (Where there is such a path, utterances like 'We met between the gate and the house', 'Let's meet between the gate and the house', 'did you see a coin I dropped between the gate and the house?' will in many contexts be taken to refer to the region of the path, by someone who knows there is a path.)

Since there is all that semantic indeterminacy/context sensitivity, that raises the question of how such linguistic expressions are understood, and in particular how we restrict those areas so as to allow utterances to be true or false, questions to have determinate answers, and commands to be definitely obeyed or not. My suggestion, elaborated below, is that we do this by treating the linguistic expression as referring to an abstract function that has to be applied to information from the context to work out what region or volume is intended: in that sense, understanding a linguistic communication often requires creative, but principled, problem

solving.

However, there is much research that starts from the assumption that there is context-free semantic content and then attempts to find out what that semantic content is is by using empirical or theoretical arguments to support a way of dividing up space, and possibly adding a changing field of values, so that within a region that could be described as between A and B, or a region that could be described as to the left of A, there is variation of goodness, or appropriateness, of locations within that region.

An example of a purely mathematical way of doing this for a region containing a collection of objects of varying shape and size is the use of a Voronoi tessellation of the region, as described in A Voronoi-based pivot representation of spatial concepts and its application to route descriptions expressed in natural language by Edwards et. al. (Proceedings of the 7th International Conference on Spatial Data Handling, 1996).

Part of the motivation for such techniques is the (mistaken) belief that in order to understand sentences referring to spatial structures, for instance route descriptions, we always have to build some sort of mental image in the form of a detailed spatial structure, from which the mistaken conclusion is deduced that machine that understands language will also need to do this. This seems to have been the motivation for the work of Olivier and Tsujii (ACL 1994) on A Computational View of the Cognitive Semantics of Spatial Expressions..

I am not disputing that spatial models are often useful, as I pointed out in 1971 in criticising the purely logicist approach to AI. But exactly what that means has to be treated with care. For instance, we sometimes use infinite spatial models, e.g. for the natural number series, as discussed here but it is clear that we cannot actually construct infinite visual images, which implies that at least some of our spatial mental structures will need to be schematic and extendable, as opposed to concrete, as first pointed out by Kant around 1781 in his Critique of Pure Reason.

Ways of using context

There is a body of work, including the aforementioned work on Voronoi-based decompositions of space, that accepts the role of context in determining what regions of space are referred to by spatial prepositional phrases, but assumes that the only relevant context is the physical distribution of objects in the situation, including possibly locations of speaker and hearer. An impressive example of this is Chapter 8 of Kelleher's A Perceptually Based Computational Framework for the Interpretation of Spatial Language in 3D Simulated Environments (2003) which not only divides up space but also computes real-numbered applicability values for each location in the space producing a sort of potential field for each region, derived from an ingeniously computed 'origin' for each region where the values peak.

This assumes that in a particular physical location where there are some occupied and some unoccupied regions of space, not only will it be possible to determine which portions of space are on the left of a particular object (once the frame of reference determining whose left has been settled), but it will also be possible to assign goodness measures to particular points in those regions. Thus the appropriateness of alternative possible behaviours of someone instructed or requested to 'Put the marble on the left of the blue cube' would be precisely graded according to which location was chosen. Likewise statements of the form 'The marble

is on the left of the blue cube' could instead of simply being true or false, also be given values that vary with the location of the marble (assuming everything else is fixed). The potential field will also determine goodness of answers to the question 'Is the marble on the left of the blue cube?' I offer an alternative theory below, which does not depend on potential fields or probability distributions.

Theories of vagueness

I think such ideas involving measures of goodness are partly inspired by experimental research on use of vague or partly indeterminate verbal expressions that ignore some of the important determinants of communication.

If you give lots of people (especially experimental subjects who are paid to produce responses) pointless instructions in a context-free experimental setting, such as 'point to the left of the blue block', or 'put the marble to the left of the blue block', or 'is the penny to the left of the blue block', then there will be some randomness in the results and the examples may be distributed in some way that suggests a potential field. But that does not mean that the semantics of the prepositional phrase includes anything to do with that potential field. It may just be an artefact of the experimental situation, for the reasons explained in the following sections.

Many people try to interpret all vagueness of semantics in terms of probability distributions (probabilities and statistics are particularly fashionable at present).

An alternative analysis of 'vague' expressions

Consider the following alternative. Suppose the semantics of a so-called 'vague' word W, (such as 'tall', 'long', 'heap', 'left of') were expressed as a higher order function, Fw, which had to be applied to features of the context of dialogue in order to determine semantic content, where the features to which the function Fw is applied include the point, or purpose of the communication, as well as physical and other facts about the scene. So the interpretation of W in a particular context, would be the result of

```
Fw(purpose, linguistic context, physical context, speaker, hearer...)
```

(Here 'purpose' need not be restricted to goals shared between speaker and hearer: during synthesis of an expression it could make use of goals of a speaker and assumed goals of the hearer, and during understanding of the expression it could make use of goals of the hearer and assumed goals of the speaker.)

The result of that function would be another function that can be applied to the environment to determine a range of acceptable heights, lengths, numbers of objects, or locations. In that case selection of a value in the relevant range of values (length, height, number of items in the heap, spatial locations, etc.) will be determined by what achieves the purpose of the communication in that context.

Example: you are trying to lay down a blanket on a beach for a picnic and the wind keeps blowing it away. Someone says, 'Let's put a big stone on each corner'. Then what counts as a big stone in that context is not, as some theories would imply, a stone that is larger than the average size of all stones (or all stones in the local vicinity), but rather something heavy

enough to keep the blanket down, but not too heavy to carry, and not something that will take up too much space, etc. I.e. it should be big enough and not much bigger. Although those limits (determined by the context, including wind strength, and shared purposes) are vague, if you randomly select one of the available stones you can decide whether it is in the appropriate range using practical judgement, without needing any precise measurements. That does not require matching to a probability distribution either.

[A similar comment could be made about 'Let's put a pile of stones on each corner', obviously.]

Likewise, if you are bringing something to me and I say 'Put it here' pointing, you don't need to compute the exact intersection of the long axis of my finger with the surface in question, then compute a potential field. If you know what the purpose of the utterance is, you can choose any location in the rough area of the surface pointed at that achieves that purpose, and assume that if it doesn't suit the speaker's requirements it will be moved, or a more specific request to move it given. Of course, when you look at what else is on the table you may be able to work out that some available locations are better than others, e.g. because of how other objects restrict what I can reach.

Note added: 29 Sep 2009
The above amounts to a solution of the Sorites paradox, which is not included in this Stanford Encyclopedia of Philosophy article on sorites: http://plato.stanford.edu/entries/sorites-paradox/

An alternative context-relative analysis of spatial prepositions

Exactly the same arguments apply to 'to the left of the red block'. If you know why being to the left is important then you can work out what range of locations will do, and choose one randomly, or on the basis of some other criterion, e.g. minimal effort. If you don't know the purpose you may try to guess or just act randomly within the region you think is acceptable, leaving it to the original speaker to suggest a change (which can be done with or without giving an explanation). Or you can ask!

If all these semantics-determining factors are excluded from a psychological experiment to find out what people do in answer to questions or commands, then the subjects will either guess what the point might be and perhaps guess differently, or perform in a random way. If that randomness produces a probability distribution that will just be a reflection of (a) the history of previous solutions to the more determined problem (which will vary from individual to individual and probably from one age group to another, and from one culture to another) and (b) the mechanisms for random generation in human brains.

I.e. the statistics will not determine the semantics, if the semantics happens to be a function of some sort, taking values from the context to compute a referent. Rather the statistics will be a very indirect product of the semantics and the peculiarities of the experimental situation, plus human willingness to comply with instructions from experiments.

One way to show all this is that if most of the space on either side of the red block is taken up, but there's an empty region somewhere on the left side (in some arbitrary direction and distance from the block), then if I ask you to put the marble to the left of the block you'll choose the empty region and there will be no reluctance about being off the 'centre' of the force field and no pressure to try to move as close to the centre of the force field within the available

region. You will not feel any need to apologise for not having space to put it anywhere near the centre of the alleged potential field. Likewise if you are asked to put it on the right, and most of the relevant region of the table is already occupied, you will not simply compute the location nearest to the centre of the force field. Rather, depending on the context, you might enter into some dialogue about whether to move something else to make space or or randomly put the object somewhere beyond the clutter, or perhaps just reply 'there's no space'. (There are many alternatives.)

The effect of qualifiers

I am not saying that there cannot be a preference ordering over locations in the region determined by the preposition. Things change if the prepositional phrase is qualified, e.g.

```
put it as near as possible to the left of the block
put it as far left as possible from the block
put it on the left of the block but not too close.
```

All of the qualifiers essentially change the function that determines the reference, sometimes in such a way as to require more inputs to be extracted from the context in order to determine the reference (e.g. 'too close for what?'), though there will very often still be a range of locations satisfying all the constraints, in which case a random selection, or a convenient selection, is all that's needed.

Qualifiers can specify multiple constraints and preferences, e.g. 'as near as possible to the block but as far as possible from the near edge'. The hearer would have to work out which near edge is referred to, using visual or other information. Likewise the assumed goals of the communication and other aspects of the context would determine the order in which to apply the constraints, which need not always be the same as the sentential order. (First find the set of *available* locations as close as possible to the block, and then order them by distance from the edge in question.)

Grice's maxims

I think all of this follows from Grice's maxims for conversational communication (which are actually relevant beyond conversational contexts) summarised here:

> http://plato.stanford.edu/entries/grice/
> http://en.wikipedia.org/wiki/Gricean_maxim

His ideas, first proposed about 50 years ago, have had wide-spread influence in philosophy, linguistics and psycholinguistics. However, I believe that they have some implications that may not have been noticed.

A brief summary of Grice's view is that understanding a communication is not just passive receipt and interpretation of a signal, like decoding morse code according to a rule book. The speaker typically has to make inferences, for example, explaining why something obvious was said, in order to work out the point of saying it. (A nice example is described in Kai von Finkel's note The puzzle of alphabetical order asking why published papers often have a note saying "The authors appear in alphabetical order", when that is obviously true and probably of no interest.)

Phone conversation example

Another example of the application of Grice's principles is using information about why something is *not* said and something *obviously irrelevant* is said instead. E.g. If you telephone someone and ask a question, like 'Is Joe with you?' and the answer seems to be totally irrelevant, like 'Yes, we were planning to sort that out at the meeting tomorrow' you may be able (depending on the situation) to infer that someone else is in the presence of the person at the other end and he wishes to give the impression to that person that he is talking to a work colleague, not a close mutual acquaintance. That requires creative problem solving by both speaker and hearer, though perhaps only the first time the device is used. (On subsequent occasions it could simply amount to use of an implicitly adopted convention, followed because it happen to work the first time it was tried.)

This is interesting in part because there are two communications in parallel, one deceptive and using something like Grice's maxims to deceive the other person in the room and one informative, using Grice's maxims to convey quite different information to the caller. (I don't know whether Grice ever considered examples like that.)

Moreover I interpret Gricean principles as allowing the hearer to understand something the speaker could not communicate because he did not have the information. E.g. A asks B on the phone 'What's Joe's phone number?' and B replies 'He said he wrote it on your note-pad'. A may then look down and see the number which answers his original question, gaining information from B's reply that B could not have given him.

Suppose *hearers* have creative problem-solving and inference capabilities that can be used to derive information from what was said, using all sorts of background knowledge about the speaker, the hearer, the physical context, some past history, or whatever. In that case, *speakers* will be able take advantage of that fact, using their creativity and problem-solving capabilities, depending on the listener also being a creative problem-solver who has access to useful relevant information. This does not assume that people who communicate have shared goals. In fact they may be antagonists in an argument, or may be trying to negotiate a deal in which each hopes to win as much as possible at the expense of the other, or one may be trying to persuade the other to leave the building.

In summary, if Grice's maxims are usable by speakers and hearers, then we can expect a language to make use of many semantic functions whose inputs are not linguistic, but depend on information being available in the environment, in the listener's mind, or in some other source the listener has access to. The non-linguistic information used as input to the processes of production and interpretation may be different and the information gained from the communication, though intended by the speaker, may be partly unknown to the speaker.

NOTE (added 19 Oct 2006)

> There is an interpretation of Grice's maxims that implies that the hearer uses problem solving or inference mechanisms to derive an interpretation of the original utterance that can always be expressed as another unambiguous utterance in the language. This is not implied by the position here. The ultimate interpretation of what was said is here assumed to be something that may intrinsically include features of the current context for which there need not be any unambiguous form of expression in the language being used.

A paper (1993) on reference semantics by J.R.J Schirra (1992), which also refers to Grice's maxims, comes close to making some of the points about higher order functions: A Contribution to Reference Semantics of Spatial Prepositions: The Visualization Problem and its Solution in VITRA. But the author is seduced by probability distributions, and wrongly takes the listener's task to be to determine which spatial (or visual) situation is 'typically' associated with the verbal description, instead of asking *which spatial situation makes most sense in the context of the verbal utterance including the assumed communicative goals*.

But he does discuss ways in which other factors can affect the typicality distribution. Still it's an interesting paper.

NOTE: Other uses of spatial prepositions

There is a different question from the one I am discussing for which the Gricean mechanisms described here provide no explanation. That is the question how users choose between alternative prepositions, e.g. 'in', 'on', 'at', 'above', 'near', 'left of', 'behind', etc. For that there may be general purpose, context-free, rules to do with spatial configurations (as described in Automatic Categorization of Spatial Prepositions by Lockwood et al., Proceedings of the 28th Annual Conference of the Cognitive Science Society. Vancouver, Canada. 2006)

Compositional semantics generalised (Added 12 Mar 2007)

The above discussion amounts to a proposal to generalise the notion of 'compositional semantics', proposed by Frege, and normally summarised something like this

The meaning of a complex expression is determined by the meanings of its parts, and the way in which those parts are combined.

For example, the semantic function (S) which derives semantic content from a syntactic structure of the form:

```
F(X, Y, Z)
```

Could be expressed as

```
S(F(X, Y, Z)) = S(F)(S(X), S(Y), S(Z))
```

For example, the arithmetic expression

```
sum(33, 99)
```

would be evaluated by applying the procedure called 'sum' to the numbers denoted by the symbols '33' and 99'.

Our discussion generalises this to take account of context and current goals (C, G) at every level, i.e.:

```
S(F(X, Y, Z)) = S(F,C,G)(S(X,C,G), S(Y,C,G), S(Z,C,G))
```

Where neither C nor G are linguistic elements, though in some of those contexts they may not be needed. For instance in many mathematical and programming contexts, such as evaluation of arithmetical expressions, C and G will not be needed. However human communication is

much more complex and they will often be needed.

Implications of Grice's theory for evolution of language

Now consider the impact of all that on the evolution of language. There are (at least) two kinds of evolution, namely (a) the evolution of the genetic makeup of humans that gives them the ability to use languages of various sorts (internal as well as external languages) and (b) the evolution of the languages they use. The latter is a form of social or cultural evolution and will typically happen much faster than evolution that depends on changes to the human genome. I am talking about the second sort of evolution given the prior evolution of basic mechanisms that make the acquisition and use of language possible (though the cultural evolution can, over time, influence further evolution of the genetic makeup, especially if linguistic communities are isolated long enough).

Suppose that there is pressure for a language to be as compact as possible (subject to many constraints) so that the information a language learner has to acquire, in order to be a competent user is no more complex than necessary. This would mean that there is pressure to increase the *productivity* of a language, i.e. the variety of types of meanings that can be expressed should increase while the complexity of the means of expression (measured in terms of the number of separate things that need to be learnt in order to express these meanings) should decrease. Obvious examples would be the use of operators such as 'and', 'or' and 'not', for combining meanings. If you understand their use, then once you have learnt to use 'That is a square', and 'That is red', you do not separately have to learn to use 'That is square and red'. If you then learn to use 'That is a block', you do not separately have to learn to use 'That is a red, square, block'. Those are examples of productivity where new meanings are composed by combining linguistic elements which already have meanings. This requires the grasp of functions (in the mathematical sense) which take meanings as inputs and produce more complex meanings as outputs. Another example is illustrated by the differences between roman and arabic numerals. The former allow some productivity, though it is very limited. The arabic notation allowed every possible positive integer to be expressed using only ten separate symbols by a simple special purpose recursive rule in which putting symbol X to the left of symbol Y means multiply X by ten to the power of the number of symbols included in Y, and add the result to Y.

All of that is old and obvious (at least since [Frege](#)) and probably earlier. But suppose we allow that people can also learn to use functions that combine meanings with other things to provide new meanings. This requires learning functions that may take in information about the physical environment, about the speaker, about the hearer, or about anything else, to be combined with meanings provided by words in a linguistic expression, to produce new meanings. That would allow the productive power of language to be even greater, since it could build on items of information that are not linguistic but have been acquired non-linguistically as part of being an intelligent agent interacting with a complex environment.

A further generalisation would be that these functions could be stacked. I,e. a function might take in some aspects of the linguistic and non-linguistic context, and produce a new function that takes in more aspects and produces a meaning, which is a statement that could be true or false, a question to be answered, an command or request to be acted on etc. In that case, getting the final interpretation would require two function applications, the second using the result of the first. Understanding an utterance could require several layers of such function

application, which implies that the language learner learns higher order functions (functions that produce functions), which need not only be second order, but could be third or fourth order, etc., or, if recursive, of arbitrarily high order.

From this viewpoint the invention of differential and integral calculus by Newton and Leibniz, and many other notational inventions in mathematics, science, engineering, and art e.g. chemical formulae, matrices, tensors, formal grammars, programming languages, musical notations, dance notations, painting and drawing conventions, were all just applications of the basic human ability to create new higher order functions that can be applied to preexisting functions to create new kinds of functions, and to invent ways indicating those functions by means of perceivable structures. The use of a 2-D surface, as opposed to a stream of vocalisations allows different sorts of complexity in perceivable structures to be used for communication, as does the use of gestures, head or body movements, etc. (Compare the case of sign language discussed below.)

If Frege's analysis of universal and existential quantifiers in ordinary language (All(Every, Each, Any, etc.) and Some(A, at least one, there is, etc.)) is correct then they would just be older examples of the sort thing. And if I am correct so would the semantics of many other words including "better", "heap", "tall", "between", "beyond", "efficient", spatial prepositions, and many more including many words previously thought to have vague meanings.

This general idea is not a new idea in linguistics, since many linguistic theorists have proposed semantic theories in which meanings are produced by the recursive application of higher order functions -- but to linguistic elements provided by the sentence being interpreted or the larger linguistic context. So all I am doing is generalising that to allow the higher order functions to take in *any kind of information*, whether acquired perceptually, by inference, from memory, or possibly even by guessing. This reduces the need for the evolution of language to be based on evolution of special *linguistic* mechanisms, since the same ability to invent and use higher order functions may serve both linguistic and non-linguistic purposes. (How it is actually done, and what sorts of mechanisms make it possible, and what sorts of factors stimulate the realisation of different applications of this competence, and how it develops in childhood, and whether any other animals have the capability in any form, are all deep problems requiring further multi-disciplinary research. There seems to be much in common between the mechanisms required for this productive ability to create information structures and the mechanisms required for fully deliberative competence, discussed here.)

> NOTE ADDED 22 Oct 2006
> It should be clear that the process of learning a language requires the use of higher-order functions which determine how the learner should use the context to determine what sort of language is being used in the environment. Moreover, what can be learnt in later stages is a product of what is learnt in early stages. This is a feature of much human learning, which is discussed in this slide presentation on interactions between genetic factors and environmental factors when learning uses genetically determined meta-competences:
> http://www.cs.bham.ac.uk/research/projects/cosy/papers/#pr0604
> 'Evolution of ontology-extension: How to explain internal and external behaviour in organisms, including processes that develop new behaviours.'(PDF)

The role of cacheing

The discussion so far has not mentioned an important point. We have seen that Gricean mechanisms, resting on general human creative problem-solving capabilities, may be very important when some form of words is either produced for the first time to express something, or understood for the first time by a hearer. However, when either the same form of words, or the same higher level pattern in a forms of words, is used or heard repeatedly, there is a mechanism that makes it unnecessary to constantly repeat these problem-solving processes whenever the same sort of need to communicate arises or a previously heard type of utterance is heard again.

It is a well known fact that humans can start off by laboriously understanding something new and complex, whether it is learning to count for the first time, learning sums and products, learning to play a game like chess or noughts-and-crosses, learning to drive a car, learning a new branch of mathematics, or learning to read music, that process can become fast and fluent because larger and larger patterns that have been experienced often enough (or in some cases only once) are stored and made available for re-use. Exactly how that works will vary from case to case, but whereas the former processes probably use neocortical mechanisms the stored re-usable patterns are likely to use older, possibly even sub-cortical, brain mechanisms. (This would correspond to differences between deliberative and reactive mechanisms in an architecture.)

This general ability depends on the fact that although in principle a complex recursive structure-generator can produce enormously complex and varied sets of structures, in fact, insofar as its use is constrained by a common environment and recurring common human needs, there will be, at least among the smaller structures, recurring patterns that can be detected as recurring and then stored for re-use. This means that sometimes Gricean mechanisms of both linguistic production and linguistic comprehension can be short-circuited by invocation of such patterns. (Essentially this idea was presented in J. Becker. The phrasal lexicon. In *Theoretical Issues in Natural Language Processing* (TINLAP 1), Cambridge, Massachusetts, 1975, available here.)

The cacheing mechanism can be used for a precise syntactic construct, e.g. interpreting 'Can you X?' or 'Would you mind Xing?' as a request to do X rather than a question about your abilities or your dislikes -- e.g. 'Can you pass the salt?' said when the salt is clearly in your reach and you obviously can pass it, as opposed to 'Can you come tomorrow?' when trying to arrange a meeting. But it can also be used for more abstract patterns of interaction like the telephone example above, where different forms of words may be used on different occasions for the same joint purpose, namely to give the person nearby the wrong impression about who is calling and to give the caller the information that now is not a good time to pursue the intended conversation.

Implications for development of linguistic competence
(Added 16 Oct 2006. Expanded 12 Dec 2006)

It may be possible to test these ideas by studying how children learn to use words with the kind of indeterminacy illustrated in the examples, including indexicals ('here', 'there', 'soon') spatial prepositional phrases, and words that refer to region range of numbers or measures ('large', 'big', 'heavy', 'pile', 'heap' etc.)

Do the children have to be presented with ranges of examples from which they induce probability distributions? (How many?, and do they induce different semantics if the frequency distributions vary from one child's environment to another?)

Or do they have to learn first how to determine what the *purpose*, or *point* of a question, statement, or imperative using those expressions is and how to use the purpose, in conjunction with other aspects of the environment, in order to work out what sort of range of possibilities will serve the purpose, and then choose from that range either on the basis of goodness for the purpose, or else on some other criterion (e.g. convenience), or else choose randomly, if there are no constraints?

Another possibility is that children start off learning to use such expressions in a nearly 'extensional', or 'prototype-based' fashion: e.g. 'big' things are only the things that adults refer to as big in the presence of children, and which 'look big' to the child: there is no higher order context sensitive function determining what is and isn't big, at first. In that case the semantics would change when they later learn to use the same expressions in the more sophisticated way described here.

Perhaps some neural deficits can impede or prevent that development, because the use of higher order functions require a more sophisticated mental architecture. I suspect it takes time to learn the appropriate generalisations, and the learning cannot occur before the architecture has grown. This paper says nothing about how that learning proceeds. The progression from prototype-based uses of concepts to higher order function-based uses would be part of a general process of developing higher order abstractions.

> (A four and a half year old child recently told me very firmly that I could not talk of a mixture of colours in the garden or a mixture of sand and sugar. Only a liquid could be a mixture, e.g. of water and fruit-juice. Presumably he will later learn what it was about the liquids that made them mixtures and apply the same formula to other cases. I don't think that treating this as metaphorical extension is the right analysis: it's moving to a higher level of abstraction.)

If we can show that *one* general mechanism suffices for a variety of different concepts and has the merit of supporting the purposes of communication, when there are purposes, that should be a better explanation than one that merely fits statistical facts about behaviour in some artificial testing situations where there is no communicative purpose (apart from understanding what to do to please the experimenter).

Question to be investigated: has anyone looked at whether the deaf children in Nicaragua mentioned below developed 'indeterminate' indexicals and spatial prepositions whose referents have to be determined by the context, or do they only use distinct determinate expressions for all the different cases? Likewise do they have words like 'large', 'small', 'thin', 'pile', or only 'larger', 'smaller', 'thinner', 'more', as well as specific names for absolute measures?

(The ideas presented here imply that the comparative concepts are more fundamental and require less sophisticated cognitive processing. It would be interesting to find out which versions came to used in human languages first. Similar questions arise concerning pre-verbal cognitive processing.)

Video:
https://www.youtube.com/watch?v=pjtioIFuNf8
This document
http://www.indiana.edu/~langacq/E105/Nicaragua.html
suggests that verbs are used instead of some of our prepositional expressions. Compare
"exceeds", "overshadows", "extends beyond".

A related question is whether probability distributions have to be learnt when children learn to use expressions referring to an amount or portion of something without determining the exact quantity, e.g. in expressions like 'a piece of string', 'a portion of meat', 'a number of people', 'a region of field', 'some distance away', 'some time later', etc.? Or do they, as suggested above first learn what sorts of purposes can be served by referring in that vague way to something, and how the acceptable values in the range are determined by the purpose of the communication?

Further implications for evolution of language

If these evolutionary pressures began from the earliest stages of the evolution of human language, that might account for some of the diversity of grammatical forms and other features of languages. It is well known that isolation of different groups within a biological species can lead to evolution of genetically distinct new species. In the case of the evolution of language, isolation would be provided by the fact that nothing like the present day means of global communication and travel were available. So as new mechanisms of expression were discovered at different times and in different orders in different linguistic communities they would spread (as 'memes') within the communities and change the context for development of new mechanisms, e.g. new higher-order functions. Some of the features of this development would be determined by shared physical and biological factors but many would be highly idiosyncratic and dependent not only on the geographical conditions but also on accidents such as who first thought of a new way of communicating something that caught on and influenced future developments in that community. In some cases historical accidents could lead to a less than optimal system that would be hard to restructure into a more effective or economical form.

It is a great pity there was nobody with relevant skills and recording apparatus to capture the processes involved in the amazing and rapid creation of an entirely new sign language by deaf children in Nicaragua. The fact that this can happen proves a number of important things. Assuming that these children were not a new species of mutants, and shared the features of human genome that enable all kinds of language learning it follows that

○ Human language learning is inherently a cooperative creative process and does not depend on the prior existence of a language used by adults that can be absorbed by observation and imitation.
(Of course where there is an existing linguistic community that will heavily constrain the learner's creativity!)
○ Human language does not depend on and is not constrained by features of human hearing.
○ Human language does not depend on our ability to vocalise and is not constrained by problems of speech production.
○ Human language is not inherently sequential, since both sign languages and many

mathematical and scientific notations are not restricted to a linear succession of atomic symbols. Deaf signers use movements of both hands, including various independent movements of fingers and changes of orientation of parts of hands, and in parallel with that, movements of head, lips, eyes, etc. I.e. several concurrent streams of spatial process are produced by the sign-language producer and perceived by the sign-reader.

> A five minute video including examples of the invented language is available here
> http://www.pbs.org/wgbh/evolution/library/07/2/l_072_04.html

> (Actually spoken languages are not simple sequential streams either, insofar as stress, intonation contours, speed variations, and accompanying gestures and facial expressions, can all help concurrently to determine what is being communicated. Moreover expert touch-typing, while it produces a sequence of characters, notoriously involves parallel development of motor patterns, one of the reasons for common typing mistakes. And although written texts are usually thought of as linear sequences of characters, reading takes in chunks that appear to be to some extent processed in parallel as happens in all vision. In any case, if the theory here is correct, both during construction and during understanding of linguistic communications the core linguistic processing depends crucially on concurrent processing of the extra-linguistic context, which can be continually changing. E.g. A asks B 'Where are the keys?' B doesn't know and starts replying 'Somewhere in this room...' then catches sight of the key-fob and continues '... oh, under that magazine on the table'.)

○ The uniqueness of human language among animal competences is probably a manifestation of something more general that is unique: the ability to create, use and interpret higher order functions, a capability that would be manifested in many different ways, including the ability to make tools to make tools to make tools, ... etc. and the ability to discover commonalities in different structures and processes at many levels of abstraction.

○ All this gives a new perspective to the influential ideas of anthropologist Lucy Suchman, (reviewed here) which had a deep impact on some strands of AI research --- causing 'situated' to become nearly as frequent as 'and' in some of the literature (as 'embodied' has been in other strands of the literature).

○ The existence of higher order functions whose arguments are non-linguistic may have started with inputs coming from the physical environment. But in principle they can come from anywhere, mathematical structures, scientific theories, religious beliefs, stories, purposes, values, ethical principles, etc. What Karl Popper referred to as the Third World (e.g. abstract cultural objects such as scientific theories, styles in art, architecture, etc. shared knowledge about many things) provides many sources of non-linguistic input. This is possibly one of the many causes of diversity in evolution of language.

○ The model proposed here, lacking in precision as it is, may also provide a useful alternative framework in which to address the concerns that have led some theorists to require all symbols (or all concepts) used by an intelligent system to be 'grounded' in sensory data, as proposed by Harnad here. In various discussion papers and presentations I have argued that this is just a reinvention of concept empiricism, a theory of the origins of meaning that goes back at least as far as philosophers like David Hume, and which was refuted by Immanuel Kant in 1780. See for example Sensorimotor vs objective contingencies and this discussion of the problem of deriving concepts of 3-D structures and motions from 2-D sensory data Requirements for going beyond sensorimotor contingencies to representing what's out there. Some of my arguments had been made long ago in What enables a machine to understand?,

(IJCAI 1995) and [Reference without causal links,](#) (ECAI 1986), and more recently in a slide presentation [Getting meaning off the ground: symbol grounding vs symbol attachment/tethering](#) attacking symbol grounding theory. That anti-concept-empiricist stance is totally consistent with the notion of meanings generated by applying higher order concepts to a combination of linguistic and non-linguistic inputs. The inputs can come from theories using indefinable theoretical terms, like [information, meaning, semantic content.](#)

○ A common thread throughout this discussion is the role played by non-linguistic context and non-linguistic competence in linguistic communication. A related point that has been made by John Barnden (and probably others) is that 'metaphor' is not an essentially linguistic phenomenon, though it is mostly discussed in the context of linguistic metaphors. Here are three pictures showing how a three year old child quite spontaneously and with evident delight used a tennis-ball and shuttlecock to create a non-linguistic metaphor:

- [picture 1](#)
- [picture 2](#)
- [picture 3](#)

John Barnden's work on metaphor can be found [here.](#)

Moreover, just as language is as much a medium for thinking as for communicating, so are metaphors, the general sense, as much a tool for thinking (e.g. understanding unfamiliar, puzzling, or complex things) as for communicating.

Where these ideas came from
Expanded: 9 Oct 2006

The idea that the meanings of indexical words and phrases should often be expressed as functions that get their inputs from the details of the context of utterance is something I got from Bonnie Webber's PhD thesis (*A Formal Approach to Discourse Anaphora.* Harvard University, 1978) where she used that idea in dealing with anaphora within discourse. I can't recall whether she also used that idea in the way I've described, for dealing with extra-linguistic reference.

It's also related to the PhD thesis of Chris Mellish in Edinburgh (*Coping with Uncertainty: Noun Phrase Interpretation and Early Semantic Analysis* 1981) which suggested dealing with ambiguities of reference by incrementally collecting constraints from the context, to narrow down the options. In his case the constraints all came from the utterance, but there's no need for that restriction.

Of course I originally learnt about the possibility of human languages as making use of higher order functions from the writings of [Frege](#). Learning about Church's Lambda Calculus and higher order AI programming languages such as Lisp and Pop-11 built on that foundation.

In the late 1960s I became concerned that work in meta-ethics, by philosophers, the meanings of words like, 'good', 'bad', 'ought', 'should', etc were missing something, and that they all depended ultimately on 'better'. However the use of 'better' seems to be so general that none of the meta-ethical theories seemed to get near explaining its function. So I came up with the idea that it was a kind of 'logical constant' a higher order function, as explained in ['How to derive "better" from "is"'](#) (American Philosophical Quarterly, Vol 6, Number 1, Jan 1969, pp 43--52). The theory was quite complex allowing for more or less elliptical uses of 'better', but attempted to define a notion formalised as

Better(P,Q,C,R,S,Z)
I.e. among things that are members of Z, things that are P satisfy condition C to a higher
degree in respect R than things that are Q in circumstances S.

This was broken down into two disjunctive alternatives, being absolutely better or being
comparatively better. The details do not matter here, but I now realise that instead of this being
a single 'flat' function there may be a collection of higher level functions, for example one
which takes a condition C and circumstances S, and returns a set of relevant respects R1, R2,
... etc. each of which corresponds to another function that can be applied to members of Z
along with C, etc. Exploring the full implications of this idea, and testing it against linguistic
phenomena, may take a mini-research project.

The idea that many words and phrases should be thought of not as directly determining
reference, but as having semantics by expressing functions which take arguments from the
linguistic and extra-linguistic context and then compute the reference, looks to me like an
important way of taking 'situatedness' into account. I suspect that this notion has not been
much investigated in this way by people working on computer-based dialogue systems, but
that may just be my limited exposure, because I spread myself so thinly.

After most of this note was written, I realised that the ideas here are closely related to this
Frege-inspired paper:

A. Sloman, 'Functions and Rogators', in J.Crossley and M.Dummett (eds), *Formal
Systems and Recursive Functions* (Proceedings Oxford Logic Colloquium, 1963), North
Holland, 1965.
Now online at http://www.cs.bham.ac.uk/research/projects/cogaff/07.html#rog

The paper makes a distinction between *functions*, which deal entirely with unchanging
mappings between arguments and values, and *rogators* (derived from the Latin for 'ask'),
which are like functions except that their values can depend on how the world is. It was
presented at a conference full of logicians and mathematicians who were completely
bemused by it, and also rightly much more impressed by important new papers presented
by Kripke, Montague and others. Clearly I am talking about rogators throughout this
paper.

The idea was originally developed in my 1962 DPhil thesis *Knowing and Understanding*,
defending Kant's philosophy of mathematics, now available online:
http://www.cs.bham.ac.uk/research/projects/cogaff/62-80.html#1962

The primacy of non-communicative language
Another old piece of work that is relevant to the ideas here, is a paper The primacy of
non-communicative language, written in 1979, arguing that the need to *represent* structured
meanings arose before the need to *communicate* them, and that this need exists, along with
mechanisms for meeting the need, in some non-linguistic animals and pre-linguistic children.

The most general linguistic construct (to be completed)
(Added: 9 Oct 2006)

An outcome of the above discussion is that human linguistic (and perhaps some non-linguistic) communication, can be seen to be made up of uses of high order functions of something like this form

```
    F1(l1, l2, l3, ... lk,  c1, c2, c3,.... cm, f1, f2, f3, ...fn) --> F2
```

Where the *li* are linguistic items, the *ci* are non-linguistic items of information, which can come from perception, memory, current intentions, or anything else, the *fi* are functions of the same general type, and *F2* is in general another function of the same form, but may at the bottom level be a truth value, or a referent (e.g. an object, or event, or process, or idea, or anything else, referred to), or possibly a method for producing a truth value (where the original was a yes-no question), or a method for determining a reference (where the original was a 'which', 'who', 'what', 'where', 'when', 'why', or 'how' question, or an instruction that can be followed, where the original was an imperative communication.

The point about Grice's maxims is that the actual words, signs, signals, produced by the communicator need not specify all the contents of the communication. Very often they will be inferrable by the hearer, or may even need to be provided by the hearer without the communicator knowing what content the hearer fills in. (This is related to the use of existential quantifiers in communications.) For example, someone says 'I am in love with someone and I have no idea what she feels about me.' His friend responds 'Try to make friends with her and talk about common interests, and then perhaps her feelings will become clear'. The friend is leaving it to the original speaker to provide a referent for 'her'.

Non-linguistic communication is covered by cases where *k = 0*.

Obviously spelling this out in detail requires much more work, and it probably overlaps a lot with existing linguistic theories, except possibly for the emphasis on non-linguistic inputs to functions that are part of linguistic semantics. See, for example, the work on Combinatory Categorial Grammar, led by Mark Steedman.

An obvious objection to everything here is that it seems to be completely inconsistent with the (alleged) success of statistical approaches to linguistics and natural language processing. The (by now, I hope, obvious) reply is that that (partial) success is explained by the same phenomena as make the cacheing, described above, useful for a language user. The statistical methods (in linguistics and AI, e.g. neural learning models) will have exactly the same general limitations as a cache in a brain, even if they attempt to combine the caches in many brains.

Links and references

I suspect other people have had the same idea about the evolutionary implications of Grice's theory, since once stated it seems so obvious (though not all the details). So I shall attempt when I have time to locate discussions of the ideas and add them here.

○ I have already found one that looks very relevant though I have not read it all yet:

Conversational Maxims and Principles of Language Planning, by Hartmut Traunmueller
In working papers PERILUS XII (1991), pp. 25-47 (Department of linguistics, Stockholm university).

There are probably many more.

○ Added: 7 Nov 2006:
http://www.mpi.nl/world/pub/BLSshort6.pdf

H.P. Grice on location on Rossel Island
By Stephen C. Levinson, Max Planck Institute for Psycholinguistics Nijmegen, the Netherlands
In S.S. Chang, L. Liaw & J. Ruppenhofer (eds.), *Berkeley Linguistics Society,* 25, 210-224. 2000
Appeared in BLS 25, 210-224

An Extract:
On this view, language codes only highly schematic and incomplete meanings. The illusion of determinate messages is due to a huge body of inference triggered by those feeble cues that constitute linguistic meaning. In short, language is sketchy. Why should that be? One crucial motivation is that human language is encumbered with a striking bottleneck in speech production: ....

○ Added 7 Nov 2006:
http://ihd.berkeley.edu/childlanguageandevolutionoflanguage.pdf
Dan I. Slobin: From ontogenesis to phylogenesis: what can child language tell us about language evolution?
To appear in J. Langer, S. T. Parker, & C. Milbrath (Eds.) (2004) *Biology and Knowledge revisited: From neurogenesis to psychogenesis.*
Mahwah, NJ: Lawrence Erlbaum Associates.

○ Added 18 Feb 2007:
Arbib, M. A. (2005)
From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics.
Behavioral and Brain Sciences, 28(2):105--124.
Preprint available here

http://www.bbsonline.org/Preprints/Arbib-05012002/

Extract from Arbib's article:
He writes

My hypothesis is that:
Language readiness evolved as a multi-modal manual/facial/vocal system with protosign (manual-based protolanguage) providing the scaffolding for protospeech (vocal-based protolanguage) to provide "neural critical mass" to allow language to emerge from protolanguage as a result of cultural innovations within the history of Homo sapiens.

The theory summarized here makes it understandable why it is as easy for a deaf child to learn a signed language as it is for a hearing child to learn a spoken language.

Comment:

I think Arbib's ideas (which I had previously read, and forgotten when I wrote most of this essay) partly overlap with the suggestions made above about evolution of language, but differ in not stressing the need for rich internal languages with compositional semantics as precursors to uses of language for communication, as hypothesized in the paper on primacy of non-communicative language. On the other hand, Arbib's work has far more neuroscientific detail, and it may turn out that a synthesis is possible.

NOTES

○ 17 Oct 2006:
It occurs to me that it would be interesting to try to extend these ideas to programming languages (if that has not been done already). That is, a programming language could include functions/procedures whose arguments can be arbitrary entities, like the current goals, current visual or other sensory information, current incomplete actions, where the results of the functions or the actions produced are not based on some simple pre-determined deterministic (or probabilistic) rule, but can depend on a creative problem-solving capability used at run time. Simple examples of this already exist in programming languages for robots or embedded controllers where conditionals can depend no the values of sensor readings, etc. But there are many more general and unconventional possibilities to explore. Perhaps that's what the so called 'autonomic' computing project of IBM should be investigating (though the people who chose the name were probably unaware that the autonomic nervous system in humans is the dumb automatic part mainly concerned with bodily functions).

10 Nov 2006
Mark Steedman has drawn my attention to debates about the semantics of 'aspect'. For example the verb 'arrive' refers to the instantaneous end state of a continuous process of motion. So we can ask when Fred arrived, and the answer will be a reference to a point of time. However we can also use the verb in ways that suggest that arriving is an enduring process, e.g. 'We got to the station as the train was arriving', 'While he was arriving I took a snapshot of him', From the viewpoint of this paper, there are many functions from a continuous process that ends discontinuously, to something else. E.g. one function takes such a process and returns a time interval that starts before the end point and ends at the end point. So 'X happened while the train was arriving' would locate X in such an interval. How to select the time interval would depend on context and the goals of the utterance. There are many such functions and many different syntactic forms for referring to them. E.g. 'X happened after the train had arrived' requires a function from a continuous interval with a discrete end to the set of time intervals after the end. There are many other such functions. E.g. 'We prepared the demo while people were arriving' depends on a function that relates a concept of discontinuously bounded process to the concept of a time interval containing a collection of end points of such processes.

- 4 Sep 2009
  The discussion here is closely related to the discussion in a document about differences between Gilbert Ryle's notion of "logical geography", concerned with the system of concepts that happens to be in use in a certain linguistic community, and a notion of "logical topography", concerned with the underlying structure of reality that supports different logical geographies, just as the architecture of matter increasingly revealed by scientific research supports different ways of classifying different kinds of stuff, some based mainly on appearance and observable behaviour, others based on the underlying physical-chemical structures. See

  http://www.cs.bham.ac.uk/research/projects/cogaff/misc/logical-geography.html

[TO BE CONTINUED]

Acknowledgements

Maintained by Aaron Sloman
School of Computer Science
The University of Birmingham