SOAS Working Papers in Linguistics Vol.10 (2000): 317-329

Meta-morphological Speculations

Sean Jensen

0. Introduction

Despite the considerable differences that exist between competing theories of phonology and competing theories of morphology, it is still possible to undertake a meaningful investigation into the nature of a language user's morphological knowledge and its acquisition in terms which are generally acknowledged to be invariant across all theories. We attempt in this paper to develop a framework which is as abstract as possible (in the sense of being independent of existing theories of particular morphological or phonological structure), yet still claims to make empirically testable claims about the acquisition of morphological knowledge as manifested by a native speaker's judgements. Far from attempting to replace current theories of phonology and morphology, this framework is orthogonal to them, providing a tool for exploring the relationship between phonology and morphology independently of any particular model of phonology or morphology.

The contribution these speculations claim to make is to make clear what phenomena can potentially be explained *outside* any theory of phonology or morphology. This sounds innocuous enough, yet the implication is that if these phenomena can indeed be explained along the lines outlined in this paper, then there is no need to build mechanisms into phonological and morphological theories which do the same work. In a slightly more imperative form of words we might assert that such theories should be 'pruned' so as to remove this duplication of effort.

This report is meant as a preliminary 'white paper', and accordingly no attempt is made here to do any of the pruning suggested by the results, although the targets of such pruning should be readily identifiable from the discussion. In particular, it could be concluded that those parts of a morphological or phonological theory which define an individual's 'grammar', as opposed to those parts which define what a general well-formed linguistic object is, are precariously overgrown.

The most surprising aspect of these speculations is the discovery of just how much can be achieved outside phonological and morphological theory, with the implication that future theory-pruning will be more than merely cosmetic topiary.

1. Objects and Forms

We shall take certain things for granted. In the first instance we assume the existence of 'phonological objects', of 'morphosyntactic objects' and 'conceptual-semantic objects'. Adopting Saussurean arbitrariness, we assume there are no necessary relations between objects of these types. Hence the objects of our study will comprise triples of these objects, which we shall call 'linguistic forms', and any such triple we shall assume is a well-formed linguistic form.

We do not need to define or even try to understand the internal structure of these three subcomponents. One of the more interesting results of this investigation is the realization of just how much linguistic knowledge and behaviour can be characterized *without* needing these internal details. In a field of study with so many competing theories and formalisms in all three sub-components, this paradigm may provide a locus for an overarching consensus.

2. Difference and Equivalence

What is required, however, for any given theory of the three subcomponents, is the ability to define the notions of *difference* and *equivalence*. It seems to be one of the most rudimentary facts of linguistic knowledge that a speaker/hearer can say whether two forms are 'the same' or not, and that these judgements of equivalence are expressible at the level of the three subcomponents. Hence we are able to say that **dɔg** and **dɔg** are phonologically equivalent, morphosyntactically equivalent and conceptual-semantically equivalent. In fact, observation reveals that our judgements are more fine-grained even than that, for example we can say that **dɔg** and **dɔt** are phonologically equivalent in the **dɔ** parts, but differ in the final part. Similarly the morphosyntactic objects associated with *dog* ('count noun, singular') and *dogs* ('count noun, plural') are equivalent in the 'count noun part', but differ in the 'singular' and 'plural' parts. And finally conceptual-semantic objects can be judged equivalent or partially equivalent: the conceptual-semantic objects associated with *dog* ('members of the set of dogs') and *big dog* (members of the intersection of the set of dogs and the set of big things) are equivalent with respect to the 'set of dogs' part, but differ otherwise.

This points to a fundamental linguistic ability: the recognition of *differences*. All linguistic theories acknowledge this by the fact that they are developed in formalisms which are essentially 'compositional' (or *algebraic*): taking smaller objects and combining them into bigger objects. If we have objects A and B in some theory, and their combination $C = A \oplus B$, the idea of the difference between C and A being B follows naturally. Equivalence is just a special case of difference, namely a difference of zero.

We therefore stipulate that whatever theories may be devised for the three subcomponents, we must additionally be able to define measures of difference with respect to those theories, such that we can take any two linguistic forms, as expressed in the calculi of the theories, and calculate the extent of their equivalence. Insofar as we can devise a difference measure in conjunction with a particular theory that can produce statements of calculated difference which correspond to an individual's actual intuitions about equivalence and difference, we have a mathematical structure whose behaviour is indistinguishable from that of a human being, and for which we could claim a measure of empirical corroboration (rather like Turing's (1950) test for artificial intelligence).

Finally we note from observation that given a linguistic difference D as just defined above, an individual's linguistic knowledge is capable of identifying pairs of linguistic forms whose difference is exactly D. For example, given the difference (*loves,loving*) we 'know' that the same difference exists between *hates* and *hating*. To the notion of difference, as introduced above, we require that there exist equivalence classes over differences and say that a *difference class* is a set of pairs of linguistic forms with equivalent differences.

A difference class, as just defined, is the generalization of the idea of 'relations' between linguistic objects. These relations are expressed in existing theories in the widest variety of formalisms, which range from traditional rewrite rules, through rule components, parametrically constrained derivations, to probabalistic and statistical rules. Although these formalisms are fiercely defended by their proponents, and disparaged by their opponents, they nevertheless all do the same thing: express relationships between linguistic objects. And from a mathematical point of view, all that means is that they define sets of pairs of linguistic objects — difference classes.

3. States

An important facet of linguistic intuition we must consider is 'acceptability', or the idea that individual language users have a sense of what sounds, or is, 'right' in their language. So we assume that an individual's linguistic knowledge is equivalent to a set of linguistic forms, some subset of the set of all possible linguistic forms, which we shall call a *linguistic state*, or simply *state*.

We are not interested in *how* this set is physically realised in the individual's brain. The answer to this question will certainly dictate the formal structure of a *particular* theory of linguistic forms, doubtless one which makes claims to psychological reality. But our interest is not in building particular theories, rather it is in uncovering those properties which must hold of all adequate theories. And at a high level of abstraction, it is true that all linguistic theories are methods for specifying what sets of linguistic forms characterize an individual's language.

In the most basic case, and that commonest in linguistic theories, linguistic forms are characterized with respect to an individual as being *grammatical* or *ungrammatical*. Thus a particular theory of grammaticality is mathematically speaking a mapping from linguistic forms into the range {*grammatical,ungrammatical*}, which means the theory can be seen as a *characteristic* of the set of grammatical forms (with respect to an individual). This is what we have called a linguistic state.

In what follows, we shall adopt a more general mapping which includes the grammatical/ungrammatical dichotomy, but also includes shades in between, widely acknowledged to be necessary for an adequate description of human linguistic intuition (notated with symbols such as '!','?','?*' *etc*, or in statistical theories with probability distributions). We shall simply say that a linguistic state is a mapping from linguistic forms into the set of real numbers.

4. Transitions

One of the focuses of linguistic research since the late 1950s has been the contemplation of just how linguistic states evolve during the lifetime of an individual. We therefore assume that an individual's linguistic knowledge (in the sense just defined, *qua* linguistic state) changes over time in response to external stimuli. In particular, we assume that at birth, an individual's linguistic knowledge is equivalent to the empty set, and that the process of language acquisition can be described as a sequence of linguistic states. This sequence necessarily implies the existence of some sort of *transition function*, which, as is widely accepted, maps one linguistic state into the next, depending in part on external learning stimuli. We shall discover that the transition function plays a central role in defining linguistic intuition, and provides most (if not all) of the functionality usually ascribed to an individual's grammar.

5. Form, Difference and Equivalence (formal)

As we have said, it is possible to talk quite generally about relatedness of linguistic structures without specifying in any great detail the specifics of their internal structures. We shall assume, however, that all linguistic theories, whether of the entire linguistic system or of various subparts of it, are mathematical structures which specify what is a well-formed linguistic object. For any such theory $\mathfrak{T}$, we shall talk of the *closure* of $\mathfrak{T}$, written $\mathfrak{T}^*$, which is simply the set of all well-formed linguistic objects as defined by the axioms and rules of $\mathfrak{T}$. We shall say equivalently that $\mathfrak{T}$ defines the *space* $\mathfrak{T}^*$, and any member of $\mathfrak{T}^*$ (*i.e.* any linguistic object well-formed with respect to $\mathfrak{T}$) is a *point* in that space.

For a given theory $\mathfrak{T}$, we define a *difference operator* '$-_{\mathfrak{T}}$' over the space $\mathfrak{T}^*$, saying that the difference between points a and b in $\mathfrak{T}^*$ is the value $b -_{\mathfrak{T}} a$, which it is up to us to define coherently with respect to $\mathfrak{T}$. A difference class in $\mathfrak{T}^*$, as defined in §2, is thus an equivalence class over pairs of forms having some given value $b -_{\mathfrak{T}} a$. We shall write $\mathrm{D}_{\mathfrak{T}}(a,b) = \{ (x,y) \mid (y -_{\mathfrak{T}} x) = (b -_{\mathfrak{T}} a)$, for all x, y in $\mathfrak{T}^*$, and given a,b in $\mathfrak{T}^* \}$. Further, we shall say that two points p and q are *equivalent with respect to* $\mathfrak{T}$ if and only if $p -_{\mathfrak{T}} q = q -_{\mathfrak{T}} p = 0$, in which case we shall write $p \sim_{\mathfrak{T}} q$. The difference class $\{ (x,y) \mid (x \sim_{\mathfrak{T}} y) \}$ we dignify with a special symbol $\mathbf{0}_{\mathfrak{T}}$. Thus we have $p \sim_{\mathfrak{T}} q$ if and only if (p,q) is in $\mathbf{0}_{\mathfrak{T}}$. We shall also say that form x *appears in* equivalence class D if and only if there is a y such that (x,y) is in D. We also note that it follows that $\mathrm{D}_{\mathfrak{T}}(w,z) = \mathrm{D}_{\mathfrak{T}}(a,b)$ for all (w,z) in $\mathrm{D}_{\mathfrak{T}}(a,b)$.

We assume, then, without loss of generality, that the various components of linguistic theories are mathematical structures whose closures define linguistic spaces. In particular, we assume that the three components of a linguistic form as defined in §1 are points in *phonological, morphosyntactic* and *conceptual-semantic spaces*. Particular linguistic theories will of course have different ways of characterizing these spaces, but that is of no immediate concern to us.

Further, we can also characterise these triples as points in the space which is the Cartesian product of the component spaces, which we might reasonably call *linguistic space*. A difference class in this space is a three-way intersection of difference classes in phonological, morphosyntactic and conceptual-semantic space. That is, every difference class $\mathrm{D}_{\mathfrak{L}}(a,b) = \mathrm{D}_{\mathfrak{S}}(a,b) \cap \mathrm{D}_{\mathfrak{M}}(a,b) \cap \mathrm{D}_{\mathfrak{p}}(a,b)$, for some $\mathrm{D}_{\mathfrak{S}}(a,b)$, $\mathrm{D}_{\mathfrak{M}}(a,b)$, $\mathrm{D}_{\mathfrak{p}}(a,b)$.

Finally, we introduce notations for equivalence in these respective sub-theories as follows: phonological equivalence (*p-equivalence*), $p \sim_{\mathfrak{p}} q$; morphosyntactic equivalence (*m-equivalence*), $p \sim_{\mathfrak{M}} q$; conceptual-semantic equivalence (*c-equivalence*), $p \sim_{\mathfrak{C}} q$; linguistic equivalence (*l-equivalence*), $p \sim_{\mathfrak{L}} q$. We note the following relation should hold between these equivalences: $p \sim_{\mathfrak{L}} q$ if and only if $p \sim_{\mathfrak{p}} q$ and $p \sim_{\mathfrak{M}} q$ and $p \sim_{\mathfrak{C}} q$.

6. Linguistic States (formal)

Keeping our discussion as general as possible, we have simply assumed that for each individual and each linguistic form there exists a number which is some measure of acceptability of that form for that individual (§3). This mapping we have called a linguistic state.

For brevity we shall call the number assigned to a linguistic form by the measure of acceptability (more correctly, by the linguistic state) the *mass* of the linguistic form. We shall assume further, without loss of generality, that masses are (real) numbers where the greater a mass, the more acceptable the linguistic form with that mass. That is, for the space $\mathfrak{T}^*$ (the set of all well-formed linguistic forms with respect to theory $\mathfrak{T}$), a linguistic state s is a mapping $s : \mathfrak{T}^* \rightarrow \mathfrak{R}$. We say that any form x such that $s(x) = 0$ is *vanishing in s*. The state $s_0(x) = 0$ for all x we call the *vanishing state*.

Analogously we say that a difference $(b -_{\mathfrak{T}} a)$ is vanishing in s if either of a or b is vanishing in s, and we define a vanishing difference class as one all of whose members are vanishing.

7. The Structure of Linguistic Space

We have introduced the ideas of a difference operator and difference classes. These classes define a *structuring* of the space from which their members come, rather like the gridlines of a map superimposing a structure onto the set of points on the map. They are, in other words, a way of 'grouping' linguistic forms. Some fundamental questions arise when we wish to claim that some such grouping or other corresponds to the linguistic intuitions of an individual. For example, do structures other than difference classes play a role in defining linguistic intuition? Or, given structures over the three sub-spaces of linguistic space, which combinations of these structures correspond to facets of linguistic knowledge and which don't, and why?

For the sake of argument, we make the strongest hypothesis, that difference is the only linguistically relevant structure over the various linguistic (sub-)spaces. When we look at linguistic space, which is the Cartesian triple product of the subspaces, we seem to find that not all possible difference classes play a role in defining linguistic intuition. We base this conclusion on the observation that whereas we can say that given the difference (*dog,dogs*), we can say that (*cat,cats*) belongs to the same difference class, we can't say, given a distance (*dog,wrote*) whether *any* difference other than (*dog,wrote*) belongs to the same difference class.

This appears to be a fundamental property of linguistic knowledge, and one that it is to be hoped can be derived as a theorem of some suitably insightful theory. Our explorations of acquisition later in this paper may help to shed some light, and may even provide the skeleton of such a theory. For the purposes of this paper, however, we simply take the facts for granted.

The property which seems to unite differences such as (*dog,wrote*) is that none of the subcomponents (phonological, morphosyntactic or conceptual-semantic) contains an equivalence, whereas we note that (*dog,dogs*) contains a conceptual-semantic equivalence. Our conjecture is therefore that linguistic space is structured by difference classes whose members are either c-, m-, or p-equivalent:

> **Conjecture A. The equivalence requirement**
> A linguistic space $\mathfrak{T}^*$ contains only those difference classes $\mathrm{D}(x,y)$ where at least one of $x \sim_{\mathfrak{p}} y$, $x \sim_{\mathfrak{M}} y$, $x \sim_{\mathfrak{S}} y$ is true.

This conjecture has an immediate corollary, and that is that these difference classes fall into 7 kinds, depending which combination of c-, m- and p-equivalences characterize their members. Given that we have been motivating our proposals on simple empirical observations of linguistic intuitions, we should expect an individual's linguistic intuitions to be able to identify all 7 kinds of difference class. It is fairly easy to show informally that these 7 kinds do indeed define axes along which linguistic knowledge seems to be organized, as examples fitting the patterns below are readily provided by untrained individuals given a suitably constructed questionnaire:

the number of non-vanishing differences in D (roughly, we prefer *flinked* because there are many more non-vanishing forms in the difference class D(*love,loved*) than any other difference class containing competitor forms.). We shall further assume that rather than just the number, it is some function of the masses of the (non-vanishing) forms appearing in the differences in D which are crucial. For instance, we do not wish to say that (*think,thinked*), for example, adds a great deal of weight to the (*love,loved*) difference class. We shall call such a function the *bias of* D *with respect to s*.

Now, since the state *s* assigns a 'measure of acceptability' or a mass to each form, we have an obvious way of characterising bias which uses both the number and masses of forms in a difference class: it is the average (mean) of all the masses of the (non-vanishing) forms appearing in the differences in D. We shall use the symbol $b_s(\mathrm{D})$ to refer to the bias of D with respect to *s* so defined. Note that the idea of bias is completely derivative: it is simply one way of looking at the information present in the linguistic state in terms of linguistic forms and masses. Thus,

Bias
For a state *s* and a difference class D, the *bias of D in s*, $b_s(\mathrm{D})$, is given by $b_s(\mathrm{D}) = (1/N)\ \Sigma_i^N\ s(a_i)$ for all non-vanishing a_i appearing in D.

Thus we have formalized the idea that difference classes which have very massive (i.e. more acceptable) forms appearing in them, and few light forms (i.e. few unacceptable forms) are systematically preferred by the transition function over those difference classes which have few massive forms, and many light forms. This seems to accord well with anecdotal evidence: new (or nonsense) forms in English almost always form a past tense in *-ed* : the D(*love,loved*) class contains many acceptable forms, and only a few unacceptable ones (like *thinked*, *swimmed* etc.), hence has a greater bias than the D(*think,thought*) class, which may have a perfectly acceptable form appearing in it, but has many more unacceptable forms (*drought*, *lought*, *shrought* etc.), reducing the overall bias of this class.

Of course we still need to define what a possible transition function is, so that we can provide a framework for analysing just how these biases can be acquired.

10. The Role of Difference and Competition in Transition

Acquisition, we have said (§4), is characterized as a sequence of linguistic states, beginning with the empty, or vanishing state s_0. Thus, given a linguistic state s_n we assume there is a mechanism for choosing a successor state s_{n+1}. This mechanism we refer to as the *transition function* T. Now, from our observation of the world we come to the conclusion that the transition from s_n to s_{n+1} is dependent not only on s_n, but on some *learning stimulus* encountered 'externally'. Assume this stimulus is a linguistic form x_n. We have then that for $n > 0$, $s_{n+1} = \mathrm{T}(s_n, x_n)$.

Now, one of the main questions of current theory is the observation that for a 'mature' linguistic state (i.e. an s_n where *n* is 'large'), the number of non-vanishing forms is huge in comparison with the number of stimuli (what Chomsky (1986) calls *Plato's Problem*). This phenomenon, and its explanation is one of the chief motivations for the universalism in Chomskian theory, and in particular the notion of Universal Grammar as a system of principles and parameters.

The transition function must therefore contain additional structure which results in more vanishing forms than just the stimulus in s_n becoming non-vanishing in the successor state s_{n+1}. We wish to explore the hypothesis that a solution to this problem does not have to lie in a *particular* linguistic theory, but rather lies in the idea of difference classes. The idea of a difference class is independent of any particular linguistic theory (although of course the particular theories will have different ways of characterizing difference classes), and the definition of the transition function we are about to develop is similarly independent.

Take a form *a*. For every difference class $\mathrm{D}_1 \ldots \mathrm{D}_n$ in which *a* appears there are forms $b_1 \ldots b_n$ such that (a,b_1) is in D_1, etc. That is, from a single form *a* we have a principled way to identify possibly many ('related') forms $b_1 \ldots b_n$, and hence we have a clue as to how the transition function might 'create' more non-vanishing forms than simply the stimulus form.

Now, it would appear that not all difference classes are equally important to the transition function, as we discussed in §7. In particular we noted that difference classes where there was conceptual-semantic equivalence and conceptual-semantic equivalence with morphosyntactic equivalence. We also assume that the transition function is constrained to consider only non-vanishing difference classes, and we shall speak of the transition function *creating* a point or form when that form is vanishing in s_n and non-vanishing in s_{n+1}.

Hence we shall base our account of the requirements of an adequate transition function around the idea that for a form *a*, there are difference classes $\mathrm{D}_1 \ldots \mathrm{D}_n$ in which *a* appears, such that $\mathrm{D}_1 \ldots \mathrm{D}_n$ are non-vanishing, and all forms (x,y) in the $\mathrm{D}_1 \ldots \mathrm{D}_n$ are conceptual-semantically equivalent, there exist forms $b_1 \ldots b_n$ such that (a,b_1) is in D_1, etc., the $b_1 \ldots b_n$ being by definition conceptual-semantically equivalent to *a*. For example, given a form *hate* and (non-vanishing) difference classes D(*love,loved*), D(*love,loving*), in which *hate* appears, we can say that there exist forms *x* and *y* such that (*hate,x*) is in D(*love,loved*) and (*hate,y*) is in D(*love,loving*) such that $hate \sim_{\mathbb{C}} x \sim_{\mathbb{C}} y$, whence, assuming appropriately defined morphosyntactic and phonological differences, we can deduce *x* = *hated* and *y* = *hating*.

Further, since transition is by definition a mapping from states to states, the identification of 'new' forms such as *hated* and *hating* from the paragraph above, is modelled by providing these forms with new masses in the successor state. Due account must also be taken of competitor forms, and hence also difference class bias. For example, if the difference class D(*take,took*) was non-vanishing, we might (depending on our particular phonological theory) need to consider in addition to *hated*, the form *hoot*, a competitor of *hated*. Clearly, the masses assigned to *hoot* and *hated* are not the same in a state which has a considerable number of non-vanishing forms appearing in D(*love,loved*).

11. Transition (formal)

The details of suitable transition functions are not necessary for the general discussion (a detailed description of a transition function can be found in Jensen 1999, Chapter 3), but certain criteria do need to be met by all transition functions: we must minimally have that $s_{n+1}(x) > s_n(x)$ for some *x* (the masses to be *excited*), and it is reasonable to assume that repeated excitation is asymptotic (that is, there comes a point when a mass has been excited so much, that additional excitation has only an tiny effect). This is reasonable because we do not expect to be able to disrupt an individual's acceptability judgements simply by repeating a malicious stimulus over and over until its mass becomes huge. By the same token, we need a corresponding notion of *inhibition*, which decreases the mass of a form, and further, that when some form is excited, certain other forms (its competitors) are inhibited. This is necessary because we could provide even more malicious stimuli, exciting every single form so much that every form becomes equivalently acceptable in the limit. By having the counterbalance of inhibition, this cannot occur, since the malicious stimuli would, in their excitation, prompt the inhibition of other forms.

Further, excitation of vanishing forms needs a little care, since as we have discussed in §9, not all difference classes are equal in the eyes of the transition function. Since bias was introduced precisely to capture this idea, and since bias was defined as an average mass, we shall simply assume that the excitation of a vanishing form results in it having assigned a mass equal to the bias of the difference class in which it appears.

So we begin with the following meta-level definitions:

Excitation and Inhibition
Excitation is a function e of masses m, where, (*i*) $e(m) > m$; (*ii*) there is a finite constant μ such that $\lim_{n\to\infty} e^n(m) = \mu$, for all $m > 0$. *Inhibition* is a function i of masses m, where (*i*) $i(m) < m$;(*ii*) $\lim_{n\to\infty} i^n(m) = 0$, for all $m \neq 0$.

With an inhibition and excitation function satisfying these conditions we are a step closer to defining the transition function, incorporating the observations of §10 above and the idea of excitation and inhibition working over competition classes discussed in the first paragraph of this section.

Additionally, we should be aware that the repeated encountering of a stimulus which is already non-vanishing should not be allowed to give rise to the cascade of forms available for excitation discussed in §9. If we did not take this into account, we would have to admit the possibility of the malicious repetition of the form *loved* being able to 'overwrite' all other past tense forms. Since we find the contrary in nature (forms such as *thought* are perfectly robust in the face of many *loved*-type forms), we are forced to assume that a non-vanishing learning stimulus is alone excited (and its competitors inhibited), without the cascade effect prompted by a vanishing stimulus.

Lastly, we need a special designated mass for the very first form ever encountered, which we shall call m_0.

We are now able to bring all these considerations together:

Transition
Transition is a function T of a state s and a stimulus r, where the *successor* state is $t = \mathrm{T}(s,r)$. (*i*) For the empty state $s = s_0$, $t(r) = m_0$; $t(z) = s_0(z) = 0$ for all $z \neq r$, and given constant $m_0 > 0$; (*iii*) for a non-empty state $s \neq \mathrm{s}_0$, and r non-vanishing in s, $t(r) = e(r)$, $t(x) = s(x)$ for all $x \neq r$; (*iii*) for a non-empty state $s \neq \mathrm{s}_0$, and r vanishing in s, for all x in all non-vanishing difference classes $\mathrm{D}_{\mathfrak{L}}(r,x)$ in which r appears and $r \sim_{\mathfrak{S}} x$, for all non-vanishing y distinct from x in the competition class $\mathrm{C}(x)$, $t(y) = i(s(y))$ and if x is non-vanishing in s then $t(x) = e(s(x))$, else if x is vanishing in s, $t(x) = b_s(\mathrm{D}_{\mathfrak{L}}(r,x))$; for all other forms $w \neq x,y$, $t(w) = s(w)$.

It is this definition which we propose can take the place of all those components of a theory which claim to characterize an individual's intuitions about the relationships between 'form' and 'meaning'. This bold claim is obviously the beginning of a new line research, rather than the conclusion of one. Some milestones on the way to that conclusion are sketched in the final section below.

12. Observations

At the macro-level, our framework allows us to analyse the acquisitional trends of linguistic states. One interesting question is Why is there not more 'irregular' morphology? Our theory allows for systems which abound in irregular morphology; yet a survey of the world's languages would seem to indicate that superficially at least, c-equivalent forms tend to have phonological forms which differ systematically, rather than randomly, from one another.

As time goes by, the more novel forms are encountered, the more existing forms in equivalent difference relations with this form will be excited at the expense of their competitors. In order to preserve an existing competing form that does not appear in a particularly prevalent difference class (an 'irregular' form, say), we should have to ensure that we encounter many exciting instances of that particular form. This seems to be precisely what is observed during the early stages of language acquisition. We note further that 'irregular' forms typically have a high 'token frequency' (many occurrences of the same form), whereas 'regular' forms have a high 'type frequency' (many occurrences of novel forms in the same difference classes).

In the vocabulary of our framework, bias acts as a record of type frequency, whereas mass acts as a record of token frequency.

Thus we may generalise that 'irregular' morphology can be sustained, *ceteris paribus*, in high frequency forms. We do *not* have to say that there is in general any pressure to 'regularise'; regularisation would only tend to happen to very low (token) frequency forms.

Further, as the device's experience broadens, fewer and fewer encountered forms will be vanishing, so that over the course of maturation, transitions involving 'type frequency effects' (cascading excitation §10) should lessen, giving way in importance to transitions involving 'token frequency effects' (excitation of the stimulus alone §10). This actually implies that it should be possible to observe 'over-regularisation' at the early stages of acquisition as 'regular' forms are reinforced with each encounter of a new form in an existing difference class, only giving way to the reinforcement of 'irregular' forms later on. Empirical evidence, such as it is, would seem to agree with this. This also explains neatly how it is possible to have 'irregular' forms with apparently low token frequency (like *weave/wove*): these forms *must* be acquired late in the acquisition process, when the reinforcements due to type-frequency are significantly scarcer. In English this is certainly the case. Most of us first learn *wove* when we are practically at school age.

Of course, given enough time, it would be quite possible to acquire or manufacture a language with more irregular than regular morphology. Although history has provided us with some plausible candidates (e.g. the Old Irish verbal system, Thurneyson 1946), our theory predicts that the situations where this is likely to occur must be very special indeed, namely a prolonged period of interaction with the linguistic environment that must extend well beyond the usual two or three years that gives us the predominance of 'regular' morphology we see in the world's languages.

Thus, our theory provides us with an interesting empirical opportunity, as it directly relates intensity and/or length of the period of acquisition with the (likelihood of) ir-/regularity in the resulting morphology. Further predictions include that on the average, the shorter/less intensive the period of acquisition, the more 'regular' the resulting language is likely to be (a study of the acquisition of Creoles in this framework should prove fruitful).

Further, if the investment of acquisitional effort in acquiring a language like Old Irish is too expensive (in some well-defined sense), we might expect that by cutting the investment (by shortening the learning period, for example), the language would mature before it had a chance to acquire all the 'irregularities'. The result would then be something like Old Irish > Middle Irish > Modern Irish, with their greatly simplified ('regular') verbal systems.

Of course what the stimulus to cut short the acquisitional period is is an interesting question, but not, I suspect, one that can be answered by linguists. Historians, sociologists and zoologists are probably better placed to understand the external pressures on organisms that determine the particular way they allocate resources to ensure their survival in a particular environment.

Our framework does imply that in the maturation of a human language, a substantial amount of monitoring of the environment is called for. This is apparently counter to the long held belief in generative linguistics that this interaction is in fact minimal, and degraded ('Plato's problem, then, is to explain how we know so much, given that the evidence available to us is so sparse' Chomsky 1986:*xxvii*). However, we should note the following: there is considerable empirical basis for the claim that there is prolonged linguistic interaction during maturation (Gallaway & Richards 1994); the type of interaction our theory requires is a very 'passive' one. We don't need to hypothesise that it is only forms addressed *to* the child that are used; as long as forms can be identified anywhere in the environment in general, this usage can be used to adjust masses in state transitions.

Further, by our hypothesis, the acquisition device is able to generate most of the forms it will ever need all by itself, especially in the early stages of the process (§10) (which accords well with the spirit of Plato's problem). Our device, however, overgenerates. We view the role of prolonged interaction as one of adjusting the masses of these forms, to see which ones are best suited to our environment ('the ones our parents like'). Again, the phenomenon of over-generalising is observed in children; and overgeneration and subsequent atrophy of a proportion of the initial population is a tried-and-tested strategy found in natural mechanisms (Holland 1992).

Finally, it is worth underlining the fact that the formal definition of our theory is completely independent of particular theories of phonology, morphosyntax and semantics, and is completely independent of any particular human language. The transition function is completely general, and therefore, we claim, able to model the acquisition of intuitions about 'form:meaning' relationships for *any* language from the appropriate linguistic environment.

We therefore conclude on the bullish note that the proposals sketched in this paper offer exciting new avenues of research including the possibility of computer simulations of the acquisition of linguistic intuitions of the sort we have been discussing.

Postscript

The ideas in this paper were presented at a SOAS Linguistics Department Seminar on 28 November 2000 under the title "Artificial Linguistic Intelligence", where they were set in a broader programmatic context. A manuscript, Jensen (M.S.), by the same title is available from the author (email *seanj@seanj.demon.co.uk*), who would welcome further discussion on all matters raised.

References

Chomsky, N. (1986). *Knowledge of language: its nature, origin, and use*. New York: Praeger.

Feigenbaum, E.A. and J. Feldman eds. (1963) [1995]. *Computers and thought*. Cambridge MA: AAAI Press and The MIT Press.

Gallaway, C. and B.J. Richards eds. (1994). *Input and interaction in language acquisition*. Cambridge: Cambridge University Press.

Holland, J.H. (1992). *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*. 2nd edition. Cambridge, MA: The MIT Press.

Jensen, S.A.G. (1999). *A computational approach to the phonology of connected speech.* Unpublished PhD dissertation. London: School of Oriental & African Studies.

Jensen, S.A.G (M.S.). Artificial Linguistic Intelligence. Available on request from the author: *seanj@seanj.demon.co.uk*

Thurneyson, R. (1946). *A grammar of Old Irish.* Revised and enlarged edition with supplement. Translated by D.A. Binchy & O. Bergin. Dublin: The Dublin Institute for Advanced Studies.

Turing, A.M. (1950). Computing machinery and intelligence. *Mind* **59**. 433–460. Reprinted in Feigenbaum & Feldman (1963) pp. 11–35.