Bilingualism: Language and Cognition 7 (1), 2004, 1–20 © 2004 Cambridge University Press DOI: 10.1017/S1366728904001178

KEYNOTE ARTICLE

Acquisition by processing: A modular perspective on language development*

JOHN TRUSCOTT
National Tsing Hua University, Taiwan
MIKE SHARWOOD SMITH
Heriot-Watt University, Edinburgh

The paper offers a model of language development, first and second, within a processing perspective. We first sketch a modular view of language, in which competence is embodied in the processing mechanisms. We then propose a novel approach to language acquisition (Acquisition by Processing Theory, or APT), in which development of the module occurs as a natural product of processing activity, without any acquisition mechanisms as such. The approach is illustrated and explicated through examples of the development of content words, derivational morphology, the functional category I with its variable features, and Case and thematic roles, as well as apparent cross-linguistic variation in processing strategies and the status of bootstrapping in the model. We then examine some possible applications to issues in second language acquisition – noticing the gap, the initial state, transfer, and the apparent limits of SLA – and finally offer a broader perspective on the model: its scope, its relations to other approaches, and its possible limits.

Language acquisition must surely provide us with a classic example of a multidisciplinary field. Its multidisciplinary character exposes it to two inherent risks. Firstly, some relevant domains might simply not yet be sufficiently developed to yield important insights. The second risk is that some relevant domains ARE sufficiently developed to impact usefully on acquisition studies but have been developed independently of one another with their own conceptual framework(s) and methodological traditions, impeding easy cross-fertilization.

The aim of our project is to build a cross-disciplinary platform which can bring together research on linguistic structure and research on general cognition, all framed within a real-time processing perspective, and which, in the process, can generate new insights of its own. We necessarily capitalize on more established research and especially on certain proposals that already have a strong interdisciplinary character. The proposed model has two parts, as it seeks to integrate, with modifications, two major theoretical positions elaborated by, respectively, Ray Jackendoff and Bernard Baars (e.g. Baars, 1988, 1997; Jackendoff, 1999, 2002). Here we focus on the first part. Even more specifically, we focus on the essential features of our main proposal, ACQUISITION BY PROCESSING THEORY (APT).

We first look specifically at the core of linguistic ability as a constellation of dedicated processors. We then consider the process by which it develops, i.e. first and second language acquisition, proposing a parsimonious approach. This seeks to explain linguistic development as much as possible in terms of on-line processing, without reference to mechanisms existing specifically for the purpose of acquisition. The discussion will conclude with a look at some applications to issues in second language acquisition (SLA).

The language module

This section sketches a modular account of language, both L1 and L2, focussing on Jackendoff's (1987, 1997, 2002) version of modularity and, within that version, Carroll's (1999, 2001) view of input processing. We also make reference to aspects of Principles and Parameters (P&P) theory (Chomsky, 1986; Haegeman, 1994) and Chomsky's (1995) minimalist idea that cross-linguistic variation regarding the module is restricted to the lexicon.

Background

There is considerable evidence that the mind is composed to a large extent of functionally specialized processors, or MODULES, though disagreement exists in many areas (see, for example, Fodor, 1983, 2000; Jackendoff, 1987, 1997, 1999, 2002; Karmiloff-Smith, 1992). We will favor

* We wish to thank Suzanne Carroll, Kevin Gregg, Michael Long, Roumyana Slabakova, Paul van Buren, Lydia White, and three anonymous reviewers for insightful comments on earlier drafts of this paper, and the Edinburgh Linguistics Circle for helpful comments in a talk we gave at the University of Edinburgh in November 2000. Thanks also to Ray Jackendoff for clarifying some aspects of his model for us.

Address for correspondence
John Truscott, Department of Foreign Languages and Literature, National Tsing Hua University, Hsinchu, Taiwan
E-mail: *truscott@mx.nthu.edu.tw*

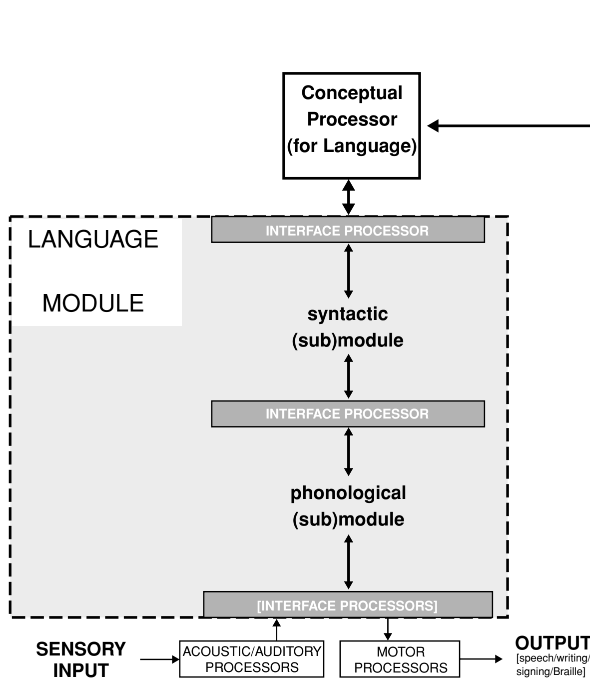

Figure 1. The language module and adjacent cognitive systems: a first sketch.

Jackendoff's conceptualization because it provides an architecture of language in a 'competence' sense that can be readily related to accounts of processing and because it is explicitly concerned with interfaces between language and other aspects of cognition.

Jackendoff's system includes INTEGRATIVE PROCESSORS, which build complex structures from the fragments of structure fed into them during processing, and INTERFACE PROCESSORS, responsible for replacing the code of one module with the corresponding code of an adjacent module in a chain accounting for phonological, syntactic, and conceptual/semantic structure (PS, SS, and CS, respectively). For example, integrative syntactic processes simply build syntactic structure within the syntactic module, using syntactic structures which are in syntactic working memory. Each integrative processor recognizes only its own code. The interface processors respond to activated (phonological or conceptual) structures outside the syntactic module and locate and activate corresponding syntactic structures for the syntactic processor to work on.

Jackendoff's theory provided the foundation for Carroll's (1999, 2001) challenging new approach to SLA. Carroll argued convincingly that a theory of acquisition must be placed within an account of language processing, and proceeded to construct such a theory for SLA, using Jackendoff's architecture. In what follows, we adopt Carroll's insight (along with Jackendoff's architecture) while differing greatly from her in its realization.

The processing chain and working memory

Figure 1 presents a simplified picture of the language module and some adjacent portions of the cognitive system. At its heart is a chain of sub-modules, responsible for processing language. We focus, here and below, on the syntactic portion. We treat the syntactic processor, chiefly responsible for building syntactic structure on-line, as invariant (see Crocker, 1996; Dekydtspotter, 2001). Morphosyntactic acquisition occurs within the lexical sub-modules and possibly the interface processors, within the constraints imposed by UG. The final product of linguistic processing is its contribution to the MESSAGE, which synthesizes the language module's output – a conceptual representation – with information from non-linguistic sources written in a generic code compatible with outputs from various domains, such as vision.

Figure 1 is elaborated in Figure 2 to incorporate our view of Jackendoff's (1987, 1997) Working Memory. It is shown, simplistically, as a two-layer space representing two levels of activation. The lower level represents a resting level and the upper represents the elements temporarily raised above that level, i.e. those in working memory. The representation is simplified in that each item actually has its own resting level and activation behavior. Working memory is often discussed as a 'blackboard', used to write (store) information temporarily during processing. But a consensus now exists among psychologists that it is not a location, as

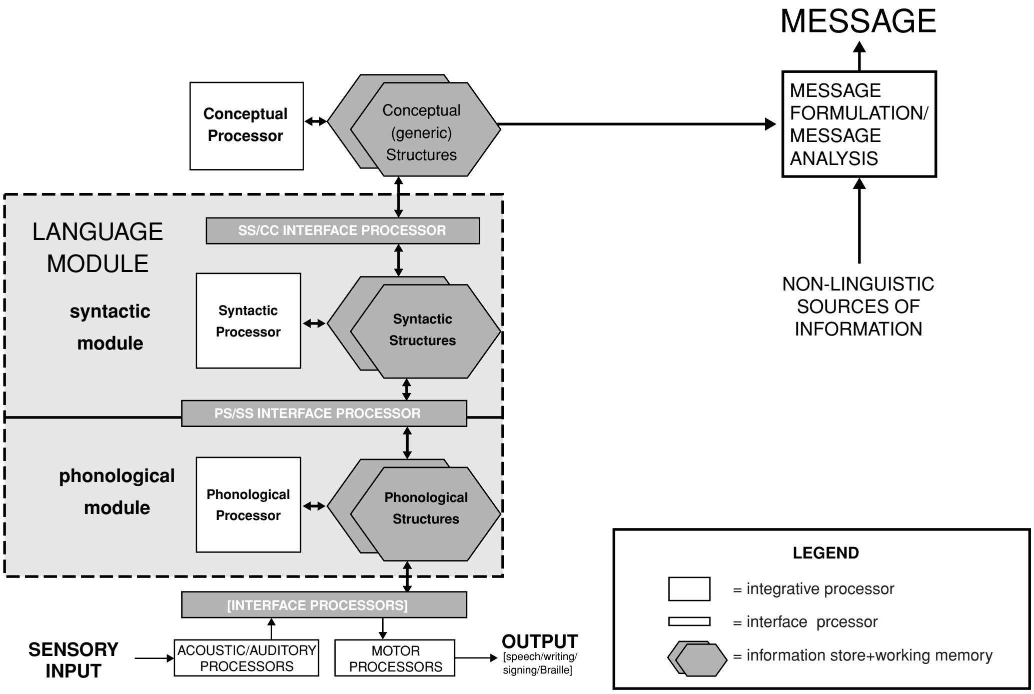

Figure 2. The language module: a second sketch with memory added.

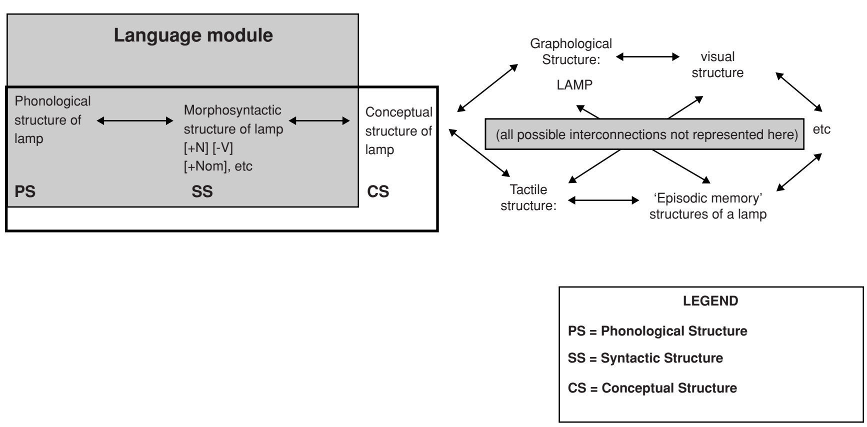

Figure 3. The modularized lexicon within and beyond the language module.

suggested by 'storage' and 'blackboard' (Miyake and Shah, 1999). It is, rather, a transient pattern of activation of elements within long-term memory stores, possibly in novel configurations. In the case of language, the long-term stores are the lexicon.

The lexicon

A lexical entry is a linking via interface processors of structures in the different sub-lexicons, as in the example of *lamp* in Figure 3, which also shows some of the many other connections outside the language module. The PS consists of the entry's phonological form, while the SS contains its syntactic category and any additional features relevant to its use by the syntactic processor. The CS represents its meaning in terms of conceptual structure.

Examples of elements in SS are features underlying a word's syntactic category and the count–mass distinction. Most important for the discussion below are the functional categories (fc's), which establish the framework for a syntactic representation (see, for example, Ouhalla, 1991; Chomsky, 1995); examples are tense, inflection, determiner, agreement, complementizer, and negation. They are language-specific instantiations of universal properties.

The conceptual (sub-)lexicon contains universal properties in the form of innate conceptual primitives. (For Jackendoff's theory, see Jackendoff, 1990.) Because the CS is written in generic conceptual code, it can be synthesized with non-linguistic information associated with the item's use. CS, though crucial to language, is thus not part of the language module; it is best seen as chunks of generic conceptual knowledge copied to a language-related store (CS) and connected to an SS and a PS. This removal of CS from the language module fits with the extensive evidence that lexical meaning has a very different status from other aspects of lexical knowledge, especially that it is acquired much more explicitly (Ellis, 1994).

The place of an L2 in the language module

SLA research has found considerable similarity with L1 acquisition and good evidence of UG availability, but also significant contrary evidence (e.g. Flynn, 1987; Bley-Vroman, 1988; Clahsen and Muysken, 1989; White, 1989; Zobl, 1990; Uziel, 1993; Poulisse, 1999). A processing-oriented framework treats all these findings as products of the processing system. If one assumes a shared system, the similarities are expected and the differences may be explained by the presence of a second set of linguistic items co-existing with and competing with the first (see below), plus the much richer metalinguistic (extra-modular) knowledge that typically accompanies SLA. If the two languages involve fundamentally distinct processing systems, differences and lack of UG availability are straightforwardly explained; but contrary evidence is troublesome. The assumption of a shared processing system also offers parsimony, allows direct application of current linguistic research to SLA, and provides a straightforward approach to incorporating the ideas of competition between languages: two knowledge bases are competing for access to a single processing system.

For these reasons, we adopt the assumption that, apart from new structures, distributed over the sub-lexicons, and possibly the addition of new lexical processes at the interfaces, the same architecture is involved in acquisition and use of an L2 as in acquisition and use of an L1. In general, our proposals therefore apply to both L1 and L2 acquisition. The distinct L1 and L2 lexical structures are clearly interconnected (e.g. Poulisse, 1999; Kroll and Tokowicz, 2001), their distinct status depending only on language-specific tagging (cf. Poulisse and Bongaerts, 1994). This conception of the bilingual mind raises interesting questions about how the L1 enters into L2 use, with important implications for transfer and other prominent issues in SLA, which we consider below.

Conclusion: The scope of 'language module'

Where, then, is 'the language module'? We will continue to speak of one, though in important respects it is not a Fodorian module but has the 'molecular' structure specified by Jackendoff, with some aspects of language use subserved by processors linked to, but not part of the module. A natural view is that it consists of those elements that are directly attributable to UG and came into existence (phylogenetically) primarily because they contribute to the function of language use. Included are the phonological and syntactic integrative processors and the interface processors linking them together and linking the module to auditory and conceptual processors. Also inside are the PS and SS stores. Excluded from the module is any grammatical, pragmatic, or other knowledge obtained by mechanisms that are not specifically linguistic, along with the linguistic conceptual processor, the message processor, and the auditory/acoustic processor(s), because they deal extensively with non-linguistic information, presumably using a generic code.

Acquisition by processing

We now offer a model of the language module's development, seeking to explain the process with minimal appeal (ideally none) to mechanisms existing specifically for the purpose; changes are to be seen as lingering effects of processing. We focus on morphosyntactic development and secondarily on conceptual development.

Processing by the language module

The process

The rich literature on processing is frequently characterized by conflicting and ambiguous findings on central issues, so every account of processing is controversial.[1] We will suggest a view which we believe is at least compatible with major findings.

One highly controversial issue is the degree and nature of interaction between syntactic and conceptual processing, views ranging from Frazier's (1979) modular approach to the unrestricted interaction of McClelland, St. John, and Taraban (1989). The logic of a hypothesized separation between syntactic and semantic processing, fitting with our modular approach, is that this specialization allows extremely efficient processing at each level. The existence of an independent syntactic processor is suggested by evidence that syntactic structures can be primed independently of semantic factors (e.g. Bock, 1986; Branigan, Pickering, Liversedge, Stewart, and Urbach, 1995) and by studies of self-paced

[1] Questions also arise about the application of reading research, which dominates this literature, to a modular approach like ours.

reading and eye movements during reading (Frazier and Rayner, 1982; Ferreira and Henderson, 1990; Rayner, Garrod and Perfetti, 1992). Research on Event Related Potentials indicates that syntactic processing and semantic processing are distinct, requiring a specifically syntactic representation (see Brown and Hagoort, 2000). But the case for some interaction cannot be dismissed (e.g. Steedman and Altmann, 1989; MacDonald, Pearlmutter, and Seidenberg, 1994).

One relatively clear finding is that semantic characteristics of a sentence can affect processing long before the end of the sentence is reached (e.g. Swinney and Hakes, 1976; Tyler and Marslen-Wilson, 1977; Traxler and Pickering, 1996). Thus, conceptual processing does not wait for syntactic analysis to be completed; results of syntactic processing are presented to conceptual processors INCREMENTALLY.

Much less clear is whether syntactic processors produce analyses serially or in parallel. Parallel processing is suggested by findings that properties of a dispreferred analysis produce priming effects (Hickok, 1993; Nicol and Pickering, 1993), indicating that this secondary analysis is present. Meng and Bader (2000), noting that experimental evidence for serial processing is disputed, produced novel evidence for it, but noted that parallel models can handle their results if they include early cutoff of unpromising analyses. There is also reason to think that the reanalysis mechanisms usually associated with serial models are necessary (Lewis, 2000), but early cutoff again allows parallel approaches to remain viable: reanalysis is the resurrection of a rejected analysis after all else fails.

Putting these themes together, a reasonable conclusion is something like the incremental-interactive theory (IIT) of Crain and Steedman (1985), Altmann and Steedman (1988), and Steedman and Altmann (1989) (see also Gorrell, 1995), adapted to the architecture described above. The syntax overgenerates, producing multiple possible representations in parallel, without reference to conceptual information. Throughout this process the SS/CS interface writes on CS whatever partial representations it can write based on what the syntactic processor has so far produced. When the conceptual processor rejects one of these CS representations, breaking the temporary connections of which it is formed and lowering the activation levels of its component items, the SS/CS interface does the same with the SS representation on which it was based. Thus, interaction between syntactic and conceptual processing consists of autonomous (over)generation by the syntax with selection among the possibilities at the conceptual level.

In adopting this leading idea of IIT, we do not adopt the details of the theory. There is reason to believe, first, that the selection process is not ENTIRELY conceptual but also involves syntactic factors (Mitchell, Corley and Garnham, 1992). The details of IIT's conceptual selection process have sometimes proven problematic as well (e.g. Hickok, 1993; Nicol and Pickering, 1993).

The need for a notion of activation level is widely accepted in processing research (e.g. McClelland and Rumelhart, 1981; Bock, 1986; Dijkstra and van Heuven, 1998). We define an item's CURRENT ACTIVATION LEVEL as its resting level plus any additional activation it has received from a processor during the current processing. Its RESTING LEVEL is determined by the extent of its past use (and possibly by innate specification in some cases; see below). Increased activation of a feature and its values follows from increased activation of the item it is associated with.

During comprehension, potentially relevant items at PS have their activation levels raised by the interface processor connecting phonological structure to (non-linguistic) auditory processing. The PS/SS interface then raises the levels of the corresponding items in SS, leading the SS/CS interface to do the same with their CS counterparts. Each type of information about the selected word is thus made available for processing at the appropriate level. Throughout, the three types are kept in registration through their common indexes.

Additional items are activated by a processor when it finds a need for them. In particular, syntactic representations require functional categories, which often have no phonological form and therefore do not receive any activation in the way just described. In a sentence such as 'They walk to work', for instance, the fc I (Inflection) must be present to satisfy UG-based conditions on well-formedness for the representation, but nothing in the phonological form of the utterance indicates its presence.

A processor uses the most active items at its level at a given moment to construct its representations. Sometimes this includes assigning indexes to empty nodes and thereby creating new items. When a processor rejects a given item or feature value, it reduces the current activation level to its original resting level. Those that are not rejected maintain their elevated levels through the current processing activity. This contrast will be central to the discussion below.

Production in the model involves the same items and mechanisms as comprehension, consistent with findings that activation of a syntactic structure in comprehension can prime it in production (Branigan, Pickering, and Cleland, 2000; Pickering, Branigan, Cleland, and Stewart, 2000). A message formulated in the code of generic structures leads to increased activation of those items in conceptual structure that could potentially be used in expressing the message through language. Once these items become especially active, the SS/CS interface processor similarly activates their coindexed counterparts in SS, leading the PS/SS interface to do the same with the coindexed items in PS. Each integrative processor works to build a legitimate representation, in its own code, by connecting some of these items along with any others it needs in order to produce an acceptable representation. In the process, it reduces the levels of items that turn out to be unsuitable for the utterance. The end result is that exactly those items that are ultimately selected will stand out in terms of activation.

An example

We will briefly present one of Jackendoff's (1997, pp. 104–105) examples of processing, with limited adjustments, as our approach largely follows his in this respect. Suppose the processing system must deal with the following sentence:

(1) It was only apparent, not real.

Assuming an American pronunciation of *apparent*, disambiguation of 'a parent' versus 'apparent' and rejection of the wrong candidate only occurs at a late stage. The phonology processor establishes both possibilities, leading the PS/SS interface to activate all items in SS that are coindexed with any of the PS items used, which then leads the SS/CS interface to do the same with the corresponding items in CS. For the case of 'a parent', the processors must also temporarily connect the two items.

The syntax processor has no problem with the input it receives from the phonology. Both candidates are well-formed even after the final fragment ('not real') is processed. However, the conceptual processor, when the 'not real' fragment is processed, requires the phrases following *only* and *not* to form a contrast. Accordingly, it rejects 'a parent', reducing the current activation of the items to resting levels. The SS/CS interface then does the same with the SS representation on which it was based. It is therefore 'apparent' that is used for construction of the message.

The lingering effects

The representations produced during processing are temporary creations, lasting only long enough to allow construction of the message. But some of their effects linger. We hypothesize two such effects.

The first involves activation levels. If an item or feature value has its current level raised by a processor and the increased level is maintained throughout the parse, the lingering effect is a small lasting increase in its resting level, the effect of which is that it becomes more readily available for future processing (because, again, processors use the currently most active items and current activation is determined in part by resting activation levels). In the preceding example, the PS, SS, and CS items that make up the word *apparent* undergo a small increase in their resting levels, but *a* and *parent* will return to their original levels; their status in the system has not changed.

The other type of lingering effect involves the addition of new items and alterations in old ones. If construction of an adequate representation for the current input requires a processor to establish a new item in SS (for example) or to alter the features of an existing item, it does so, purely for the purpose of processing the current input. If these changes are not undone during the parse and the representation in which they appear is not rejected, they remain in SS afterward. But if they are rejected during the parse, they quickly decay, meaning that no change (no learning) occurs. Again, we hypothesize this rejection as a normal means of removing inappropriate candidates from the current parse; it is not a learning mechanism.

Jackendoff (1997) argued that a rejection process of this sort must be part of normal processing. In the example just considered, the phonology processor itself cannot distinguish *apparent* from *a parent*; the distinction can only be made at the conceptual level. But the phonology processor does end up with only *apparent*, as shown by the fact that we subjectively hear one word, not two. It follows that rejection messages are passed from CS to PS, suggesting that *a parent* has been removed. A possible alternative (less appealing because it fits poorly with IIT) is that acceptance messages are sent down, further raising the current levels in the analysis that was not rejected, making this analysis stand out from those that were rejected.

Evidence regarding removal of SS candidates is more difficult to obtain, but the claim appears testable. If priming is the persistence of high current levels after a parse is completed, syntactic items present in a rejected analysis but not in the accepted analysis should not show priming effects that last beyond the current parse, because their activation levels will have been lowered by this point. This prediction might be tested with the procedure used by Branigan, Pickering, and Cleland (2000), involving dialogue between a subject and a confederate (perhaps in a written version). If the confederate uses the sentence 'The floor was uncovered wood', the subject's syntactic processor should construct, along with the correct analysis, an alternative involving passive, which should be quickly rejected. Because the processor has removed the passive (see below for our account of passive), this sentence should not make the participant more inclined to produce passive sentences afterward; syntactic priming should not occur. If there is no removal mechanism, priming should occur.

As production involves essentially the same representation-building process as comprehension, except in reverse, it should also result in lasting changes, though only of the first type. The items used in production are established through comprehension, which provides the necessary information. Thus the long-term effect of production is to selectively raise resting levels of already existing items and their feature values.

Content words

We illustrate these ideas, first, by looking at the acquisition of content words. Suppose the learner encounters the utterance 'The horse is beautiful' and the language system does not yet have any items corresponding to *horse*. Focussing on the higher levels of the processing chain, we will simply assume that the phonology is able to parse its input into words, in this case including one that did not exist prior to the parse and therefore gives rise to a new PS item, corresponding to *horse*.

The creation of this new item leads the PS/SS interface processor to establish a counterpart for it in SS, by coindexing it with an initially empty SS item. The interface also activates the SS items corresponding to the other words in the sentence, resulting in the proto-representation Art–?–V–A on SS, with each item bearing an index that ties it to the corresponding PS (and CS) items. To construct a syntactic representation for this input, the syntactic processor must assign a category to the new item and integrate it with the other items without violating UG constraints (embodied in the processor), and respecting the order of items in the proto-representation.

Assuming that some basic syntactic development has already occurred, the processor should assign the category N to '?' because doing so allows the creation of a representation in which the article is paired with a following N and a complete NP subject is present. The processor thus raises the activation level of the new item's [+N] feature and lowers that of [+V], resulting in lasting changes in their resting levels. The changes are small, and therefore subject to reversal if future input requires it for construction of successful representations. But for the case of *horse* future input will generally result in the same category assignment, further raising the resting level of [+N] and lowering that of [+V], i.e. strengthening its status as a noun.

As the syntactic representation is being produced, the SS/CS interface uses it to establish a proto-conceptual representation on CS, including an empty item coindexed with the PS and SS for *horse*. Using this information, the conceptual processor must produce a complete conceptual representation for the sentence, which requires it to assign the new item a meaning. The likely source is contextual information, which is not available at SS but is at CS. If the speaker is pointing to or looking at a horse, for example, this information could allow the appropriate chunks of conceptual knowledge to be copied onto CS and coindexed with the new PS/SS, establishing the CS for *horse*.

Other syntactic and conceptual features of a word can develop in much the same way. An input sentence such as 'Pat hit Chris', for example, should lead the syntactic processor not only to strengthen the V status of *hit* (raise [+V] and lower [+N]), but also to assign it a [transitive] feature or, if one is already present, raise its current level and therefore its resting level (assuming that neither the feature nor the representation in which it appears is rejected during the parse). The reason is simply that these features are necessary for an adequate syntactic representation for the input. Similarly, the conceptual processor can only construct an acceptable representation if it assigns *hit* a suitable theta grid, specifying the thematic roles of its accompanying arguments. In this case the grid must include a hitter (agent) and a hittee (patient). So the processor writes these features on the CS for *hit*. We return to thematic roles below.

We have so far abstracted away from the parallel character of processing. The syntactic processor routinely constructs additional representations for its input, in parallel. It might, for example, try to construct a representation for 'The horse is beautiful' in which *horse* is an adverb, like *really*, and the subject N is non-overt, along the lines of '(These creatures are ugly, but) those [$_N$e] really are beautiful'. But typically all these attempts fail to produce fully acceptable representations at either SS or CS (as in the example of 'a parent').

Derivational morphology

When the system encounters *unhappy* for the first time, the phonology module produces two alternative representations in parallel, one treating it as a single unanalyzed novel word, /ənhæpi/, the other as the novel item /ən/ plus the existing /hæpi/ (assuming that *happy* has already been acquired). For each case, it establishes a new entry for the novel item, each of which then acquires an initially empty SS and CS.

The fate of the two new entries depends on which representation survives, which depends on how successful the syntax processor is in producing from it an acceptable syntactic representation and how successful the conceptual processor is in producing from that representation an acceptable conceptual representation. The *un*+*happy* analysis has an advantage in that it can make use of the information already contained in the entry for *happy*, at all three levels, which should contribute to the success of the representations in which it appears. If the ultimate set of representations uses the *un* entry but not the *unhappy* entry, the latter fades, either disappearing entirely or remaining with a very low resting level, because the processing system rejected it. The *un* entry remains, with an initially low resting level, which future use should raise considerably.

This example can be contrasted with the processing of *until* or *unto*. Here a decompositional analysis is unlikely to produce successful conceptual representations, so only the holistic forms should survive. Thus, semantically transparent words, such as *unhappy*, are much more likely to be stored in decomposed form than more opaque words, as found by Marslen-Wilson, Tyler, Waksler, and Older (1994).

Nothing prevents the system from establishing both analyzed and unanalyzed forms, especially when the complex form is encountered first and the stem form is only acquired later, so this account predicts the presence of some redundant storage, as in dual-route models (e.g. Laudanna and Burani, 1985; Chialant and Caramazza, 1995). It also predicts that experience with a derivational affix will make the affix more available for future processing; frequency of the affix should thus be associated with its availability in processing.

The SS and CS for *un* (and possibly for *unhappy*) develop in essentially the same way as those for content words. The CS can come from non-linguistic information: learners who hear *unhappy* in context can often judge that its meaning is the opposite of *happy* and that *un* therefore means 'the opposite of', allowing that chunk of conceptual knowledge to be copied into CS and coindexed with the appropriate PS and SS. Assignment of features to the SS – containing the information that it is a prefix and attaches to adjectives – results from characteristics of the morphosyntactic context. In order to construct an acceptable representation involving *unhappy*, the syntactic processor must write these features on the *un* SS. They will then remain after the parse is complete, with initially low resting levels, which will be raised whenever future processing uses the features.

Conclusion

This approach does without some familiar concepts related to acquisition. This is the primary respect in which we diverge from Carroll's (1999, 2001) approach. The notion of Language Acquisition Device (LAD) has no independent place in the model but is rather an abstraction, reifying various processes involved in the development of the language module. Similarly, we have no notion of processing mechanisms turning things over to separate learning mechanisms when they cannot do their job ('failure-driven' acquisition). In our model there are no learning mechanisms as such in the language module, only the lingering effects of processing within an innately constructed performance system.

Acquisition by processing: Functional categories

The heart of syntactic variation is the set of fc's provided by UG, each with its own variable features. In discussing their development, we first present an overview of the process and then examine two examples: the category I, along with two parameters associated with it, and then Case and thematic roles.

In general

The essence of an fc is an innately specified SS, pre-existing in the sense that the syntactic processor, as an instantiation of UG principles, writes it on SS when dealing with appropriate input. The set of variable features and possible values for them pre-exists in the same sense – the processor by nature writes on the fc's in SS those needed to make the current parse succeed.

The CS is similarly inherent in the conceptual processor and the SS/CS interface. When an fc is written on SS, the interface then writes a coindexed item on CS and the conceptual processor assigns it whatever features and values are needed for the current parse. These features and values are probably best seen as semantic primitives, or combinations of them, which can be copied into CS to serve as meanings for fc's (cf. Slobin's, 1985, 'grammaticizable notions'). The possibilities are strongly constrained by UG, though not so strongly as to rule out variation. The English tense system, for example, does not use the [future] feature (i.e. it is not copied into the portion of CS coindexed with the tense SS), while Spanish does.

If the syntactic processor is to write a particular category on SS, it must recognize that it needs that category, with a particular feature value, to handle its current input. But without considerable syntactic and lexical context this may be impossible. So acquisition of some content words, including their syntactic characteristics, must precede development of fc's.

But once suitable contexts can be constructed, fc's should become relatively easy to deal with. The processor is set to put them into its representations, so when its input contains the appropriate clues, determined by UG, it does so. Similarly, construction of syntactic representations requires activation of an appropriate feature value of the fc, and the syntactic processor is designed to determine which of the limited possibilities best fits with the representation it is building. Increases in the resting level of the SS and CS items come from their continuing use by the processors in constructing representations.

Establishment of an fc's PS is quite different, resembling that of content words. A given functional CS can have several distinct PSs (e.g. English past tense can be realized as [d], [t], or [ɪd]), each initially entered into the lexicon independently of the others. A new PS can be connected to the SS (and CSs) when the PS/SS interface registers the correspondence between the two in the phonological and syntactic representations constructed during a parse and coindexes them accordingly. Once again, this change occurs as part of normal processing; no learning mechanisms as such are involved.

Inflection (I)

To illustrate these points, we now consider in more detail the development of the fc I in English.

The item

I is innately present in that the syntactic processor by nature writes it on SS whenever possible. But in the early stages of learning the processor cannot determine its relevance to the given input, so it is not included in syntactic representations. We follow Hawkins (2001) in hypothesizing that the situation changes with the acquisition of the copula *be*, which the syntactic processor can recognize as occupying I. The logic is that the copula represents the most basic element associated with I, in the sense that it is a free morpheme and an expletive verb and has minimal selectional restrictions, and so should be the clearest and therefore the first clue to the presence of I. This expectation is supported by L2 findings (especially studies by Andersen and Stauble, described by Hawkins) that other aspects of I (auxiliary *be*, tense markers, agreement) develop after copula *be*. Evidence that it does indeed trigger the inclusion of I in representations is provided by the observation that as soon as learners begin to use the copula they tend to place it to the left of negation, typically avoiding such mistakes as '*He not is hungry' (studies by Stauble and Shapira cited by Hawkins). This observation suggests that the copula is in the I position rather than the V position (see below for related discussion), from approximately the time it first appears.

Because the meaning of a copula is purely relational, it can only be identified in context; in other words it must be surrounded by content words that have already been acquired. Suppose a learner has encountered the sentence 'The glass is empty'. By assumption, the lexicon already contains fully developed entries for each of the words except *be*. The phonology module is therefore capable of parsing the string into four words, with the third word initially identified only as something new.

Thus, the new PS /ɪz/ is established. The phonology-syntax interface processor then writes a new item in SS, coindexed with it, triggering the SS/CS interface processor to write another new item in CS, also coindexed. At the syntactic level, the SS pattern Art–N–?–A is initially written, where '?' represents the new item. The syntax module then produces any syntactically acceptable representations it can from this incomplete information. At least one of these is likely to treat the new element as I, because valid representations will result. This entire process is guided by UG, as instantiated in the syntactic processors.

The conceptual processor is likely to accept the representation, because doing so should allow construction of an acceptable CS. I is the natural home for a copula, so when the syntax places I in the representation the conceptual processor can (perhaps must) give the coindexed item in CS a meaning consistent with a copula. The resulting representation as a whole should be acceptable to the conceptual processor, so the SS representation identifying '?' as I is not rejected. I then remains in SS as a (potentially) permanent item. Similarly, the meaning given the corresponding CS item remains. The syntactic and conceptual items will be coindexed with one another and with the PS /ɪz/.

There are no guarantees in this process. For any given input, the syntax processor might not produce a syntactically acceptable representation that treats the new item as I; if it does the conceptual processor might not accept this representation. Thus, initial establishment of I might occur only after multiple encounters with potentially useful input.

The strength feature

Once I exists in SS, its features can be established in future parsing. Consider the [strong] vs. [weak] feature, which determines whether verbs can appear in the I position rather than in their canonical position in the VP.[2] If it is set at [strong] (the activation level of the latter is higher than that of [weak]), when it becomes part of a representation this setting will have the consequences shown in (2).

(2) *We finished quickly our meal.

Placement of the verb *finished* in the I position, to the left of the adverb *quickly*, results in a sentence that is ungrammatical in English (in contrast to French, for example). Thus, English learners must end up with the [weak] setting, which blocks this structure.

One type of input that could establish this feature value (or further strengthen it) is the appearance of a negator, *not* in the case of English, preceding a finite thematic verb, as in 'We did not finish our meal'. This situation is inconsistent with a [strong] setting, so the representation constructed by the syntactic processor must include the [weak] value for I, or else an incoherent representation will result. So the processor increases the current activation level of [weak], with small lasting effects on its resting level. A step has thus been taken toward firmly establishing that value. Additional input will have the same effect, further consolidating the appropriate setting.

One complication is that there may be an initial, default value; learners assume that one particular value is correct until they encounter evidence for the other. Platzack's (1996) argument that [weak] is the default translates in our model into a claim that when I is first written on SS, the resting level of [weak] is considerably higher than that of its alternative. It is therefore used in production until enough contrary input has been processed to raise the [strong] value to a level at which it takes over.

[2] In more familiar terms, it determines whether V can move to the I position, one of the parameters of P&P; see Lasnik (1999).

Pro-drop

The development of I's features can also account for the pro-drop parameter. P&P accounts of the parameter hypothesize an empty category, pro, which can appear in the subject position only if it is licensed. The licenser has typically been associated with Inflection; we thus assume that it is a feature of I, with the possible values [+] (directly or indirectly licensing pro, permitting pro-drop) and [−] (disallowing pro-drop).

In comprehension, when the syntactic processor constructs a representation lacking an overt subject (the phonology provides nothing that can be analyzed as subject), it seeks to place pro in that position. This is only possible if the I feature is [+], so it raises this value's current level, with lasting effects on its resting level if the representation is not rejected. The [−] value is temporarily raised when the syntax processor constructs a representation containing an overt pronoun as subject; traces of the change remain if this representation is not rejected. In production, the processor favors the value with the higher resting level, selecting pro or an overt pronoun accordingly. The choice further strengthens that value.

English speakers reject null subject sentences because the [−] value in English is very high and the [+] value very low as a result of repeated long-term exposure to sentences with pronoun subjects and very infrequent exposure to subjectless sentences. This situation makes it virtually impossible for the syntax processor to temporarily raise the [+] value above the [−], a requirement for constructing representations with null subjects. To use a handy metaphor, it would be a very costly process, given the great disparity in resting levels, so the processor avoids it. The situation is reversed in pro-drop languages: [+] has the higher level, so subjectless sentences are the norm.

Case and thematic roles

We assume Case and thematic roles are innately present in the sense that they are embodied in the processors; i.e., the syntactic processor by nature establishes items representing Case on SS and the SS/CS interface by nature establishes on CS conceptual items representing thematic roles and coindexing them with the Case items.

Case

In P&P overt Case marking is an expression of underlying abstract Case, which is always assigned to every NP (see Chomsky, 1986). The particular Case (nominative, accusative, etc.) is defined by the head that is in a position to assign it to the NP in the representation. An NP governed by I, for example, receives nominative Case.

We will assume that each abstract Case is instantiated by an fc in SS. In constructing its representations, the syntactic processor always connects each NP to one of these Case items. When no suitable item yet exists (mainly in the early stages of acquisition), it writes Case_i onto an empty node and temporarily connects it to the NP as part of the construction of its representation. The new Case is defined by the Case-marker that governs the NP it is connected to in the representation being constructed; e.g. nom when the governor is I and acc when it is V (see Figure 4 below). This new fc, created specifically for the purpose of building the current representation, will remain (assuming the conceptual processor does not reject the representation in which it appears), with an initially low resting level, which should rise quickly through continued use.

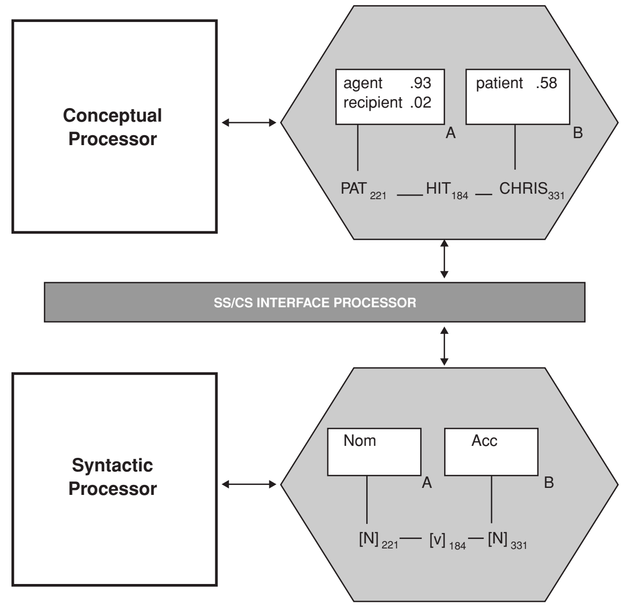

Figure 4. Acquisition by Processing at the SS/CS interface: thematic roles.

In this way the inventory of Case items in SS is established relatively quickly, on the basis of innate guidance from UG. The development of their phonetic and conceptual counterparts is more complex. We will not explore the relations between Case items and PS items here, except to note that they can be quite complex and variable, including null PSs, many-to-many mappings, and interactions with other SS items, especially those responsible for gender and number agreement.

Thematic roles

Whenever the syntactic processor writes a new Case item on SS, the SS/CS interface processor writes a corresponding item in CS, coindexed with it and therefore co-activated with it during processing; otherwise construction of a CS representation compatible with that on SS will not be possible. Crucially, these CS items do not represent thematic roles as such, but are probabilistically associated with them. Thematic roles are features written by the conceptual processor on them, each with its own activation level, which is strengthened whenever it is used in a representation. Figure 4 offers an illustration.

When a learner encounters 'Pat hit Chris', the syntactic processor is likely to assign nominative Case to the first NP (based on government by I), leading the SS/CS interface processor to activate the CS item coindexed with nominative and connect it to the CS counterpart of 'Pat'. Establishment of an acceptable semantic representation of this sentence will require the conceptual processor to assign the agent role to this item (assuming that context and/or background knowledge make it clear that Pat is doing the hitting). Thus, it raises the current level of the [agent] feature of this CS item (or writes the feature there if it is not already present), with the long-term effect of slightly raising its resting level. Through many such processing events, the [agent] feature gradually comes to dominate this CS item. Because the item is coindexed with the nominative item in SS, a strong association is thus formed between agency and nominative Case.

But this CS item is not exclusively associated with the agent role. A sentence such as 'Pat received a gift' will require the conceptual processor to write the feature [recipient] on it (see Figure 4). The two thematic roles it bears will then each constitute a distinct feature of the item, each with its own independent activation level.

Complications are introduced by passive constructions. At CS, the essence of passive is a reversal of connections. Returning to Figure 4, if the utterance is 'Pat was hit by Chris', the conceptual processor must connect PAT to the CS theta item that would otherwise be connected to the item following HIT. The cue for this reversal is the presence in the CS representation of an fc that serves exactly this purpose. It is the CS counterpart of the 'passive' item at SS (realized at PS as the verbal forms associated with passive; *be*–V–*en* in English) and is therefore coindexed with it. At SS its effect is to deny Case to the NP in the object position and thereby force it to appear in the subject position, where it can receive Case. This category is innately present in the same sense that other fc's are, i.e. the syntactic processor by nature writes it onto SS in the appropriate (pre-specified) conditions and the SS/CS interface responds by writing its conceptual counterpart in CS.

When a passive utterance is received before this item is first written in SS and CS (i.e. before passive has been acquired), the syntactic processor will begin to produce a straightforward (active) representation for it, without using a passive item, simply because it CAN quickly and efficiently produce one that is reasonably successful, without doing anything new. If extra-modular knowledge is available, indicating that Pat is the hittee, the conceptual processor will reject this analysis before it is completed, cutting short this line of activity by the syntactic processor. At the same time, the syntactic processor seeks to produce other representations in parallel, possibly including some that require the creation of new items or major changes in existing ones (though such analyses are probably subject to abrupt termination by the SS/CS interface when a more conservative analysis proves acceptable to the conceptual processor). One of these possibilities is to write the passive item on SS, leading the SS/CS interface to write its CS counterpart. If this option is pursued within whatever time/space constraints may exist (during this parse or on some future input), the conceptual processor will not reject the resulting representation (because it leads to a successful CS representation), so the passive item will remain in SS, coindexed with the appropriate item in CS (and PS). Afterward, it will be easier to get a passive interpretation for future passive input because the necessary items are already in SS and the processor will not have to put them there.

Initially low activation levels limit this ease, but they should rise with future input. Increases could also come from production, which cannot establish new items or features in SS but can strengthen existing ones. Once the items/features are present and have coindexed counterparts in CS, if the conceptual processor activates the latter in production then the associated SS items will also be activated and possibly have their resting levels raised as a result.

APT and the competition model

This approach can incorporate some important insights underlying the Competition Model (Bates and MacWhinney, 1987; MacWhinney, 1987, 2001). The focus of research within that model has been on how the agent role is assigned to the appropriate NP, cross-linguistically, using competition among a number of cues, including semantics, word order, Case, agreement, and passive morphology. Speakers must learn to weight these cues on the basis of how useful each is within their language. In English, for instance, word order is dominant because it is a very reliable cue (the agent is the NP preceding the verb, except when passive morphology is present) and is consistently available. In Italian, word order is more variable and therefore less important; agreement takes center stage because it is the best cue. Quite generally, the strategies speakers acquire for assigning the agent role are products of the formal characteristics of the individual languages, as in these examples.

The APT approach hypothesizes no language-specific processing strategies as such. Instead, cross-linguistic contrasts are a product of acquired differences in the content of the lexical stores, primarily the fc's. We will illustrate the approach by discussing the way that word order and agreement influence assignment of the agent role in English and Italian.

For this purpose, we adopt the possibly simplistic view that the subject's position in a sentence is determined by direction of Case assignment. The fc I, which assigns nominative Case, has a direction feature with the values [left] and [right], each with its own activation level, influencing the direction in which I assigns Case and therefore where the subject NP must be in order to receive Case. In English the [left] value has an extremely high level while the [right] value is virtually zero. Thus, I must assign nominative to the NP to its left; in other words, when the syntactic processor constructs its representations, the 'nominative' fc is consistently attached to the NP in this position.

When the SS/CS interface processor begins its work on the CS representation, it connects the two corresponding elements in CS, namely the cluster of thematic values that is coindexed with the nominative item in SS and the conceptual structure representing the NP to the left of I. Because nominative has become strongly associated with the agent role, through experience, the [agent] feature has a far higher activation level than any other features in the cluster and so the conceptual processor is pushed to interpret the NP it is attached to as agent. In most cases, this interpretation fits well with the verb's theta grid. In 'Pat hit Chris', for example, the grid of *hit* contains an agent slot, which is naturally filled by *Pat* because the syntax processor ties it to the nominative item, which is coindexed with the CS item in which [agent] is the dominant feature.

A non-agent interpretation of the first NP would require a temporary reversal of the activation levels in the thematic cluster, raising another feature (say [recipient]) to a level higher than that of [agent]. This would be a very costly move, given the great disparity in resting levels, so the processor would avoid making it unless forced to. What would force it to is the presence of a verb with a theta grid that does not include an agent. An example is *receive*, which requires a recipient and not an agent. Thus, the dominance of the [left] value of I and the acquired association between nominative and agent virtually force the conceptual processor to interpret the NP to the left of the verb (to the left of I) as agent, exceptions occurring when they are forced by the meaning of a particular verb.

In Italian, with its more flexible word order, this enormous disparity between the levels of [left] and [right] never develops. The nominative SS item can be attached to an NP in either direction, depending on other factors, especially agreement.

Following P&P accounts, we treat agreement as an fc in SS. It is coindexed with its own thematic cluster in CS. If the agreement is between subject and verb (as opposed to object and verb), the [agent] feature will have by far the highest activation level of the roles that make up the CS cluster. Association of the agreement element with a particular NP at SS therefore results in a similar association of agent with the NP at CS, just as in the case of nominative Case assignment. The conceptual processor is thus pushed to assign the agent role to this NP. Italian has rich overt subject agreement, which is used with great consistency. The agreement fc therefore acquires an extremely high resting level, as does its CS counterpart and the latter's dominant feature, [agent]. Thus, the NP to which the agreement fc is attached is almost inevitably interpreted as agent, unless (again) the verb's theta grid rules out this interpretation.

In English, agreement is not as strong a cue as word order because it is not as consistently available and therefore does not acquire the extreme resting activation levels associated with the latter. Thus, when conflicts occur between the two cues, word order dominates, because the resting activation level of the [left] feature of I is so high that it cannot be reversed even when agreement clearly calls for such a reversal. In a nonsentence such as 'The men loves the woman', the conceptual processor must still assign the agent role to 'the men', because of its position. The agreement feature must therefore be treated as an error; i.e., the sentence receives an interpretation but is judged unacceptable.

Bootstrapping

Bootstrapping (e.g. Pinker, 1984, 1987; Weissenborn and Höhle, 2001) is generally considered an essential ingredient for language acquisition, so we will explore its status within APT. In its most general form it refers to the learner's use of one type of linguistic information for acquisition of another type. Within our model, this means information at one level of the chain being used in a way that produces lasting changes in the lexical store at another level. This can in principle occur in either direction, upward or downward. Interface processors, mediating relations between levels, are at the heart of this process. Crucially, APT requires that the information be used entirely for processing purposes; the system is not trying to make inferences about characteristics of the language, only to produce representations of the current input.

Upward bootstrapping is quite natural, due to the assumption (see above) that new items and features are established specifically in comprehension. Comprehension involves the interface taking rich information at one level and using it to help build a representation at the next level up. This process opens the door to the substantial upward bootstrapping that apparently does occur, as in the way prosody influences syntactic development (see Morgan and Demuth, 1996; Weissenborn and Höhle, 2001). The most basic example is that when novel words are written in PS, the PS/SS interface can then use them to establish corresponding items in SS.

Downward bootstrapping should be more constrained, based on informational encapsulation and the assumption that new items and features are not written during production. Ideally, in comprehension (where new items/ features are written) the effects of conceptual information on syntax (SS) are limited to rejection messages passing from CS to SS. This restrictive notion of encapsulation still allows important indirect conceptual-to-syntactic bootstrapping.

In comprehension, the syntax overgenerates representations, producing all the syntactically acceptable ones it can within existing constraints, the conceptual processor issuing on-line rejections of those that do not lead to acceptable CS representations. If a new item is necessary for successful interpretation of the input and just one of the syntactic representations contains it, the conceptual processor is likely to reject all the alternatives, leaving that one representation active and thereby allowing the novel item it contains to become a permanent part of SS.

The discussion of passive above provides an example. The passive item is latently present, in that the UG-based syntactic processor is prepared to write it on SS under appropriate circumstances. But before passive has been acquired, the syntactic processor cannot 'know', in any sense, that English has a passive construction or that a passive representation is appropriate for 'Chris was hit by Pat'. And there is no purely syntactic way for it to find out. But if it does produce a passive representation as one of its parallel attempts to analyze this input, the conceptual processor can accept this representation (by rejecting all the alternatives) and thereby ensure that the passive item acquires a place in SS. Here the combination of UG-guided syntactic overgeneration and semantic selection allows the emergence of an SS item that might never be realized otherwise, except through a direct form of downward bootstrapping that would raise serious questions about informational encapsulation.

The idea of indirect downward bootstrapping is probably applicable to many additional bootstrapping cases. The possible influence of semantic cues on the determination of a word's syntactic category (central for Pinker, 1984; limited for Elliott and Wexler, 1986) might be explained by the conceptual processor accepting representations in which the syntax has given it a semantically appropriate category and rejecting those in which it has not. If learners are especially able and/or especially inclined to analyze situations in terms of an agent acting on a patient (the 'Manipulative Activity Scene' of Slobin, 1985), the conceptual processor will be more likely to accept syntactic representations that are easily reconciled with this scene, allowing the beginnings of a syntactic Case system in SS (though in a very different way from that hypothesized by Slobin).

Some applications to issues in SLA

The model has implications for many issues in SLA. We now consider four examples.

Noticing the gap

Acquisition necessarily involves a comparison, in some form, between characteristics of input and current characteristics of the mental grammar. This requirement has appeared in many forms, including Schmidt and Frota's (1986) NOTICING THE GAP. It is usually stated in a largely pre-theoretical form, with no attempt to specify how the comparison is made. Our account provides a theoretical context while abandoning the idea of a direct comparison. When the success of the current parse depends on the system having certain characteristics, the processors impose those characteristics on it during the parse and thereby push it toward a state that fits better with the current input. There is no mechanism as such that identifies contrasts between grammar and input. Instead, the comparison is an abstraction from characteristics of processing. A more subtle version of this sort is perhaps inevitable, given Carroll's (1999) observation that acquisition is based not on inherent characteristics of the signal but on characteristics of the representations the processing system constructs from that signal.

The initial state

The initial state of the L2 has been the subject of debate (see Herschensohn, 2000; Hawkins, 2001) as to whether it consists of all relevant characteristics of the L1 (Schwartz and Sprouse, 1996), its lexical characteristics but not its fc's (Vainikka and Young-Scholten, 1996a, b), or fc's with their values unspecified (Eubank, 1996). Within our processing-oriented, quasi-minimalist approach, the issue is the characteristics of the L2 lexicon and how processors use them in constructing representations.

It is commonly assumed that the presence of L1 characteristics in L2 performance (production, comprehension, judgments) implies their presence in the L2 grammar, i.e. in the L2 lexicon. But this assumption is open to question. The processing chain can access both lexicons and clearly does so during L2 production, as shown by code-switching. Furthermore, evidence suggests that lexical access is non-selective (see Dijkstra, de Bruijn, Schriefers and ten Brinke, 2000; Dijkstra and van Heuven, 2002; but see Costa and Caramazza, 1999, for some qualifications). When one lexical entry is accessed there is a varying degree of activation for a host of competing candidates in all available language systems.

So in any given instance, the appearance of L1 characteristics in L2 performance could in principle be explained in either of two ways: (a) the L1 properties have been transferred to the L2 lexicon; or (b) they are coming directly from the L1 lexicon. Given the shared processing chain and the pervasiveness of code-switching, (b) is surely the proper explanation for some cases.

One might ask, then, whether ALL cases of apparent transfer could be instances of (b). This idea suggests a radical alternative view of the L2 initial state: it contains no L1 characteristics. This NO TRANSFER/FULL ACCESS approach has considerable a priori appeal. Because the processing chain clearly does access the L1 lexicon during L2 production, in code-switching, an adequate account of linguistic performance must explain how this alternating access occurs. This appears, a priori, to be the most natural place to look for the source of L1 characteristics in L2 performance. Such an approach is appealing because it can potentially account for transfer effects using independently motivated mechanisms.

Adoption of this approach would not resolve (or dissolve) the disagreement over the initial state. It would, rather, reformulate the debate. Instead of asking what L1 characteristics are in the initial state of the L2, one would want to know under what conditions the L1 lexicon enters into the processing, production, and judgment of L2 utterances, and in what ways. Given the apparent lack of progress toward a resolution of the debate, such a reformulation could have considerable value, especially as it appeals to a research area not previously considered in this context.

Transfer

The proper interpretation of apparent transfer effects was a prominent issue in the preceding discussion. We now consider this issue in more depth, returning to the example of the strength feature of I.

White (1991) observed that French-speaking English learners acted as if English I was strong, as it is in French, producing ungrammatical sentences like (2), repeated here as (3).

(3) *We finished quickly our meal.

In an attempt to remedy the problem, learners were given explicit instruction in the correct order, resulting in short-term improvements. Schwartz and Gubala-Ryzak (1992) showed that the treatment did not change the parameter value; learners' performance continued to show a [strong] setting for I. In a related study, Trahey and White (1993) gave learners a flood of sentences exemplifying the correct English order. They found that learners subsequently accepted and produced more sentences with this order but did not abandon ungrammatical sentences like (3), indicating that once again the parameter setting had not changed. Thus, the L1 I feature appears to have transferred to the L2 and shown itself very resistant to change.

In our approach, these findings are explained in terms of competition between the two I's for inclusion in the representations constructed in SS during processing. The L1 I, I_F, has an extremely high resting level, due to constant use over a long period. It therefore has an enormous advantage over its L2 counterpart, I_E, and routinely wins the competition, explaining the appearance of sentences like (3).

The L2 I can enter when features of its L1 counterpart cannot be reconciled with the rest of the representation. When the input requires a weak I (e.g. when *not* occurs before a finite verb), use of I_F is virtually impossible because its [strong] setting has such a high resting level. In such cases I_E, with its low resting level, can win the competition and appear in the ultimate representation of the input, because it has no competitor. Recurring cases of this sort gradually raise the resting level of I_E and its [weak] value. This is, however, the only way the level can be raised. In production, there are no such factors forcing the use of I_E, so the dominance of I_F should persist, perhaps indefinitely.

In this context, the findings of Trahey and White (1993) are expected. Given the extreme resting level of I_F, even if I_E could sometimes enter the representations during the flood, it would have to do so a truly vast number of times before its resting level could reach a point at which it could seriously compete with I_F. The flood – and the explicit instruction – could influence performance only through the development of extramodular metalinguistic knowledge, with which learners could deliberately modify their output. (See Sharwood Smith, in press, for discussion of extramodular metalinguistic knowledge.) It could make learners metalinguistically aware that English adverbs can sometimes precede verbs but could not alter feature values inside the module and therefore could not affect the use of ungrammatical sentences like (3). In this context, the fact that explicit instruction was somewhat more effective than the implicit flood is also expected: explicit learning of a relatively simple point should be more effective than highly implicit learning when no guidance from UG is available.

This discussion suggests, again, that the appearance of L1 characteristics in L2 use, even when chronic and long-term, need not indicate transfer, as it is normally understood. The problem could well lie not in L2 competence but in the extent to which it can be expressed. This is, in effect, chronic involuntary code-switching.

The limits of SLA

Second language learners rarely achieve the level of success consistently reached by L1 learners. Many possible explanations are available, some of which are particularly interesting within our approach.

An input competition view

Problems for L2 acquisition could come from L1 involvement in L2 processing. During the parsing of an L2 utterance, L1 entries are also activated and therefore compete with the appropriate L2 entries for a place in the representations being constructed. If an L2 entry loses the competition then any changes it underwent during the current parse are lost; no learning occurs. If L2 items routinely lose out to their L1 counterparts, then L2 acquisition will be slow, or even non-existent.

An interesting corollary is that L1 items used in processing an L2 utterance might undergo changes to fit the input. The result could be small but lasting changes in those items. One such experience would have only a tiny effect, but repeated experiences could produce meaningful changes, particularly in an L1 item with a low resting level. So this discussion suggests that L2 characteristics can influence the L1, though such effects should typically be limited. Available evidence supports this prediction (Kecskes and Papp, 2000; Cook, 2003).

Limits in working memory

Another possible source of problems for older learners is a decline in working memory capacity. Working memory, we said previously, consists of transient activation patterns in the lexicon. In our model, a necessary condition for long-term changes is sustained activation of the appropriate items in the lexicon during processing. If the capacity to sustain activation (working memory) declines with age, language learning ability might decline with it.

The nature of working memory is a complex issue, dictating caution in discussions of 'decline' in working memory capacity. After all, children with limited working memory capacity are efficient acquirers. Too much information in working memory at the same time may complicate acquisition. A slower rate of processing may lead to a reduction in efficiency of inhibitory mechanisms, or vice versa (Salthouse and Meinz, 1995; Miyake and Shah, 1999). The issue is apparently not how much information can be held in memory at one time but how long it can be held there – the capacity to sustain activation.

An output competition view

The 'no transfer/full access' view of SLA sketched above offers another possible explanation for the limits of L2 acquisition. The hypothesis it suggests, in its strongest form, is that L2 acquisition has exactly the same success as L1 acquisition, up to the limits imposed by available input, but the resulting L2 competence cannot be fully expressed due to competition from the L1 lexicon.

Production, like comprehension, involves competition between the two lexicons for access to the processing chain (see de Bot, 1992; Poulisse and Bongaerts, 1994; Hermans, Bongaerts, de Bot, and Schreuder, 1998; Kroll and Tokowicz, 2001). Given the strong connections L1 items have already formed with the chain (i.e. their high resting levels), the L2 system inevitably faces an uphill battle, frequently losing the competition and failing to fully express L2 competence. The L2 lexicon may also have to compete with L2 knowledge outside the language module, which, given the general learning ability of the typical L2 learner, may be highly developed and may have attained strong access to output systems.

This discussion has implications for a current debate in SLA, recently extended to language attrition (e.g. Polinsky, 1997; Toribio, 2000; Sharwood Smith, 2001; White, 2001). The frequent absence of overt inflections in L2 performance suggests to some that the abstract features they reflect are absent. To others, the fact that syntax is behaving as though these features were present suggests it is a matter of 'spell-out'. Either the overt realizations of the abstract features have not been learned yet or the learner's lexicon does make them available but something happens in processing such that they fail to find their target.

This something could be output competition. Items from both languages are activated during parsing (see above). If bilinguals have difficulty inhibiting irrelevant candidates this must also stretch to parts of words, e.g. inflectional affixes, possibly explaining problems in their use. For a language attriter using L1, the resting level for L1 items is lower than that associated with a perhaps less developed L2, possibly leading to the L2 item winning out against the targeted L1 item. And, in a Jackendoff-style model, the lexical entries are triples kept in registration while separate processing of each member is carried out. Under such circumstances chunks written in the same code but from different triples might interact or inhibit each other in ways that result in an affix getting suppressed en route.

In addition to attrition and the apparent limits of L2 acquisition, the output-competition approach has potential as an account of some speech errors, transfer effects, and code-switching. However, the idea is clearly speculative at this point, awaiting more serious development. And it does not rule out other views of the limits of L2 acquisition, which might come, for example, from a combination of output competition and transfer of L1 features to the L2 lexicon.

The model in perspective

We now step back from the details to look at the model in the context of existing approaches to language processing and acquisition (first and second) and then to consider its possible limits.

The scope of the model

Our proposal is very ambitious. Its primary goal is to explain language acquisition, first and second, in a novel and theoretically appealing way. Because it roughly equates acquisition with processing, it is necessarily an account of processing as well, both production and comprehension. In a very limited sense, it is also a model of language, as we place competence within this processing framework, but our ideas here are largely drawn from P&P. We have obviously not yet developed our proposal to the same extent as models aimed specifically at explaining one particular area (e.g. Crocker, 1996, for comprehension; Levelt, 1989, for L1 production; de Bot, 1992, for bilingual production; Pinker, 1984, or Slobin, 1985, for L1 acquisition; Herschensohn, 2000, or Hawkins, 2001, for L2 acquisition; Chomsky, 1995, for L1 competence).

Relations to other approaches

A major inspiration for the model was Carroll's (1999) insight that input and acquisition should be seen in the context of processing. We are in a sense exploring the limits of this idea, considering the extent to which processing and acquisition can be equated. Our approach also resembles Carroll's (1999, 2001) in its adoption of Jackendoff's (1987, 1997, 2002) architecture of the language faculty. With this architecture comes the assumption of modularity and rich innate structure, largely as hypothesized by Jackendoff. It thus differs fundamentally from approaches such as those of Slobin (1985) or Bates, Bretherton, and Snyder (1988), which replace innate linguistic knowledge with universal mechanisms for acquiring the knowledge.

UG plays a central role in the model as a blueprint for the language module, providing the overall architecture and the details of the processors. Because the items in SS and PS are written there by these processors, UG strongly constrains these items, and to some extent those in CS, as the latter are partially determined by the UG-based SS/CS interface and SS items, especially in the case of fc's. Crucially for syntax, UG provides the pool of fc's, with their possible features and any default values. This information is built into the processors, which therefore write it in SS as soon as they determine that it is needed for the representation of some particular input.

We diverge from Carroll and Jackendoff in using P&P research, because it is a rich source of well-developed accounts of syntactic phenomena and such accounts are necessary for determining the details of SS and the syntactic processors. Minimalism contributes the idea that cross-linguistic variation in competence is restricted to the lexicon, as is acquisition. This assumption establishes clear connections to Herschensohn (2000), Hawkins (2001), and Juffs (in press), among others. We differ from previous P&P and minimalist models primarily in adopting a processing-oriented approach. This move also suggests interesting alternative perspectives on central issues within these models, as seen above.

The combining of two different approaches – P&P with Jackendoff's functional architecture – raises questions about compatibility. In most respects there is little cause for concern, as we take Jackendoff's general architecture (which assumes little about the nature of syntax beyond standard, widely shared generative assumptions) as the framework and use P&P to fill in the details at the syntactic level. The two differ most fundamentally in the interface between syntax and conceptual structure. Specifically, we follow Jackendoff in doing without P&P's distinct level of Logical Form (LF). He places LF phenomena, scopal phenomena in particular, in the SS/CS interface. Jackendoff (1997) argued in detail that these phenomena cannot be adequately handled below the conceptual level, with the implication that doing away with LF as a distinct level actually allows better explanations of LF phenomena. The use of P&P syntax without the LF component does not appear problematic either, as LF was already removed from syntax; the purely syntactic processes operate independently of LF, their output feeding into it. So dropping LF from P&P leaves the syntax unchanged and has syntactic output feeding into CS, which now includes all LF functions. Thus differences regarding LF do not appear to pose any problems for our mixing and matching approach.

We have in a sense gone beyond minimalism, applying its idea that variation is purely lexical to language processing strategies as well. Research inspired by the Competition Model, described above, shows that speakers of different languages process input differently (see MacWhinney and Bates, 1989; MacWhinney, 2001; and references cited there). According to conventional thinking, this shows the presence of distinct processing strategies – in our terms, that the processors are different. But this need not be the case. The behavior of an information-processing system is the product of the characteristics of the processor and those of the database it works with, in this case, the lexical stores. Differences in the behavior of two systems can in principle be attributed to differences in the processors, the databases, or a combination of the two. One cannot assume that the processors must differ from one another. We propose instead to place the differences in the database, i.e. the lexical stores. This move resonates with recent findings that lexical storage may be far more extensive than traditionally believed (Nooteboom, Weerman, and Wijnen, 2002).

It also has conceptual advantages. MacWhinney and Bates emphasized that the differing strategies result from differing characteristics of the languages; speakers rely on the cues that are most useful in their language. The Competition Model posits two distinct types of entities, formal characteristics and processing strategies that use them; each must be acquired. A preferable position, a priori, is that only the differing formal characteristics of the languages must be acquired; the behavior of the systems follows directly. This is a goal of APT. We cannot claim to have fully achieved it here, but we believe we have shown that the APT approach should be pursued as a promising alternative to standard formulations.

This discussion shows the potential benefits of merging minimalist ideas with a processing-oriented approach to acquisition. The most novel and important aspect of our approach, though, is its abandonment of acquisition mechanisms in the language module. Chomsky has repeatedly suggested that the development of language is not learning but rather growth, analogous to the development of physical organs. There is considerable tension between this analogy and the hypothesis of a LAD. The latter, with its deductions and learning principles, has dominated discussions of acquisition within the generative perspective, but its activities do not resemble the growth of an organ. APT offers a potential resolution of this tension, doing away with the LAD and moving acquisition much closer to the organic development idea.

The approach is primarily symbolic, using traditional processors (UG-based) which implement principles of grammar, using those principles to write and manipulate symbols in information stores. These characteristics give it several strengths: a route around poverty of the stimulus problems, a direct account of the systematicity of language, and the possibility of one-shot learning (a UG-based processor writes a new symbol on an information store).

It has clear affinities with connectionism, though, especially in that development involves changes in individual strengths resulting from success in processing (cf. the interactive activation approach of McClelland and Rumelhart, 1981). This characteristic gives it some of the strengths of connectionist models. It directly predicts the existence of frequency effects (e.g. Schwartz and Terrell, 1983; Gershkoff-Stowe, 2002) and the typically gradual, non-discrete character of learning. It explains learners' ability to recover from errors (e.g. Bowerman, 1988) – one feature value might become stronger early on, based on limited input, but subsequent input can raise another value enough to reverse the system's behavior. In this way it can also explain how learning can succeed despite the noisy data learners routinely encounter (Truscott and Wexler, 1989; Valian, 1990), providing a means of incorporating the concept of ROBUSTNESS (e.g. Lightfoot, 1991) in a primarily symbolic approach. Robustness involves repeated activation of items, resulting in lasting increases in their resting levels.

But again we do not hypothesize learning mechanisms as such, even in the connectionist sense. When a processor raises an item's current activation level as part of processing and the level is maintained through the parse, that item will thereafter have a slightly higher resting level than previously. This might be seen in terms of connectionist-type learning mechanisms, selectively rewarding components of the system that have proven useful. But a preferable view, we would argue, is that this process of change is inherent in the nature of activation levels, perhaps comparable to the way a muscle grows through exercise: one could say the muscle grew stronger because a learning mechanism selectively rewarded it for participating in the exercise, but such a description would be at best unhelpful. In adopting this view we diverge from all explicit theories we are familiar with.

Limits of APT?

The claim that development of the language module can be explained this way is admittedly quite strong, especially regarding several areas we have deliberately neglected, mainly for reasons of space, including the phonology processor and the PS/SS and SS/CS interfaces. Some issues we have discussed require more in-depth treatment, and additional issues must also be addressed, such as the establishment of syntactic categories (see Culicover, 1999) and the possible role of distributional analysis. Processing strategies do not appear to pose any problem of principle, but caution is also required here.

Language acquisition also involves the considerable linguistic knowledge found outside the module, which might be acquired differently. This category includes metalinguistic knowledge of all sorts plus orthography and other aspects of reading/writing skills. Word meanings occupy a marginal position in relation to the module, and a full account of their development should reflect this position. APT may have considerable relevance even here, though, as seen in the discussion of content words above.

Conclusion

Faced with the frustrating fragmentation within this multidisciplinary field, our goal in this paper has been to present a reasonably detailed and explicit model of bilingual language acquisition, incorporating plausible accounts of the nature of language and language use. Jackendoff's processing-friendly proposals were adopted rather than an approach in which linguistic competence issues are explained purely in abstracto, although such accounts are certainly not dismissed as irrelevant.

The main thrust has been to provide a common framework within which one can productively study many important, but normally disparate issues in the field together. These issues include – in addition to processing and acquisition – transfer, code-switching, language attrition, the role of UG and its limits, and the limits of SLA relative to L1 acquisition. Possible approaches have been touched upon here and will be explored in greater detail elsewhere.

References

Altmann, G. & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition*, 30, 191–238.

Baars, B. (1988). *A cognitive theory of consciousness*. New York: Cambridge University Press.

Baars, B. (1997). *In the theater of consciousness: The workspace of the mind*. New York: Oxford University Press.

Bates, E., Bretherton, I. & Snyder, L. (1988). *From first words to grammar: Individual differences and dissociable mechanisms*. Cambridge: Cambridge University Press.

Bates, E. & MacWhinney, B. (1987). Competition, variation, and language learning. In B. MacWhinney (ed.), *Mechanisms of language acquisition*, pp. 157–193. Hillsdale, NJ: Erlbaum.

Bley-Vroman, R. (1988). The fundamental character of foreign language learning. In W. Rutherford & M. Sharwood Smith (eds.), *Grammar and second language teaching: A book of readings*, pp. 19–30. Boston: Newbury.

Bock, K. J. (1986). Syntactic persistence in language production. *Cognitive Psychology*, 18, 335–387.

Bowerman, M. (1988). The 'No negative evidence' problem: How do children avoid constructing an overly general grammar? In J. A. Hawkins (ed.), *Explaining language universals*, pp. 73–101. Oxford: Blackwell.

Branigan, H. P., Pickering, M. J. & Cleland, A. A. (2000). Syntactic co-ordination in dialogue. *Cognition*, 75, B13–B25.

Branigan, H. P., Pickering, M. J., Liversedge, S. P., Stewart, A. J. & Urbach, T. P. (1995). Syntactic priming: Investigating the mental representation of language. *Journal of Psycholinguistic Research*, 24, 489–506.

Brown, C. & Hagoort, P. (2000). On the electrophysiology of language comprehension: Implications for the human language system. In M. W. Crocker, M. J. Pickering & C. Clifton, Jr. (eds.), *Architectures and mechanisms for language processing*, pp. 213–237. Cambridge: Cambridge University Press.

Carroll, S. E. (1999). Putting 'input' in its proper place. *Second Language Research*, 15, 337–388.

Carroll, S. E. (2001). *Input and evidence: The raw material of second language acquisition*. Amsterdam: Benjamins.

Chialant, D. & Caramazza, A. (1995). Where is morphology and how is it processed? The case of written word recognition. In L. B. Feldman (ed.), *Morphological aspects of language processing*, pp. 55–76. Hillsdale, NJ: Erlbaum.

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

Chomsky, N. (1995). *The minimalist program*. Cambridge, MA: MIT Press.

Clahsen, H. & Muysken, P. (1989). The UG paradox in L2 acquisition. *Second Language Research*, 5, 1–29.

Cook, V. (ed.) (2003). *Effects of the second language on the first*. Clevedon: Multilingual Matters.

Costa, A. & Caramazza, A. (1999). Is lexical selection in bilingual speech production language-specific? Further evidence from Spanish–English and English–Spanish bilinguals. *Bilingualism: Language and Cognition*, 2, 231–244.

Crain, S. & Steedman, M. (1985). On not being led up the garden path: The use of context by the psychological syntax processor. In D. R. Dowty, L. Karttunen & A. Zwicky (eds.), *Natural language parsing: Psychological, computational, and theoretical perspectives*, pp. 320–358. Cambridge: Cambridge University Press.

Crocker, M. W. (1996). *Computational psycholinguistics: An interdisciplinary approach to the study of language*. Dordrecht: Kluwer.

Culicover, P. W. (1999). *Syntactic nuts: Hard cases, syntactic theory, and language acquisition*. Oxford: Oxford University Press.

De Bot, K. (1992). A bilingual production model: Levelt's 'Speaking' model. *Applied Linguistics*, 13, 1–24.

Dekydtspotter, L. (2001). The universal parser and interlanguage: Domain-specific mental organization in the comprehension of *combien* interrogatives in English–French interlanguage. *Second Language Research*, 17, 91–143.

Dijkstra, T., de Bruijn, E., Schriefers, H. & ten Brinke, S. (2000). More on interlingual homograph recognition: Language intermixing versus explicitness of instruction. *Bilingualism: Language and Cognition*, 3, 69–78.

Dijkstra, T. & van Heuven, W. J. B. (1998). The BIA model and bilingual word recognition. In J. Grainger & A. M. Jacobs (eds.), *Localist connectionist approaches to human cognition*, pp. 189–225. Mahwah, NJ: Erlbaum.

Dijkstra, T. & van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition*, 5, 175–197.

Elliott, W. N. & Wexler, K. (1986). A principle theory of categorial acquisition. In S. Berman, J.-W. Choe & J. McDonough (eds.), *Proceedings of NELS 16*, pp. 136–153. Amherst, MA: GLSA.

Ellis, N. C. (1994). Vocabulary acquisition: The implicit ins and outs of explicit cognitive mediation. In N. C. Ellis (ed.), *Implicit and explicit learning of languages*, pp. 211–282. London: Academic Press.

Eubank, L. (1996). Negation in early German–English interlanguage: More Valueless Features in the L2 initial state. *Second Language Research*, 12, 73–106.

Ferreira, F. & Henderson, J. M. (1990). Use of verb information in syntactic parsing: Evidence from eye movements and word-by-word self-paced reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16, 555–568.

Flynn, S. (1987). *A parameter-setting model of L2 acquisition: Experimental studies in anaphora*. Dordrecht: Reidel.

Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology*. Cambridge, MA: MIT Press.

Fodor, J. A. (2000). *The mind doesn't work that way: The scope and limits of computational psychology*. Cambridge, MA: MIT Press.

Frazier, L. (1979). *On comprehending sentences: Syntactic parsing strategies*. Bloomington, IN: Indiana University Linguistics Club.

Frazier, L. & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, 14, 178–210.

Gershkoff-Stowe, L. (2002). Object-naming, vocabulary growth, and the development of word retrieval abilities. *Journal of Memory and Language*, 46, 665–687.

Gorrell, P. (1995). *Syntax and parsing*. Cambridge: Cambridge University Press.

Haegeman, L. (1994). *Introduction to government and binding theory* (2nd edn.). Oxford: Blackwell.

Hawkins, R. (2001). *Second language syntax: A generative introduction*. Oxford: Blackwell.

Hermans, D., Bongaerts, T., de Bot, K. & Schreuder, R. (1998). Producing words in a foreign language: Can speakers prevent interference from their first language? *Bilingualism: Language and Cognition*, 1, 213–229.

Herschensohn, J. (2000). *The second time around: Minimalism and second language acquisition*. Amsterdam: Benjamins.

Hickok, G. (1993). Parallel parsing: Evidence from reactivation in garden-path sentences. *Journal of Psycholinguistic Research*, 22, 239–250.

Jackendoff, R. (1987). *Consciousness and the computational mind*. Cambridge, MA: MIT Press.

Jackendoff, R. (1990). *Semantic structures*. Cambridge, MA: MIT Press.

Jackendoff, R. (1997). *The architecture of the language faculty*. Cambridge, MA: MIT Press.

Jackendoff, R. (1999). The representational structures of the language faculty and their interactions. In C. Brown & P. Hagoort (eds.), *The neurocognition of language*, pp. 37–70. Oxford: Oxford University Press.

Jackendoff, R. (2002). *Foundations of language*. Oxford: Oxford University Press.

Juffs, A. (in press). Prolegomena to a theory of representation and processing of a second language. *Transactions of the Philological Society*, 102.

Karmiloff-Smith, A. (1992). *Beyond modularity: A developmental perspective on cognitive science*. Cambridge, MA: MIT Press.

Kecskes, I. & Papp, T. (2000). *Foreign language and mother tongue*. Mahwah, NJ: Erlbaum.

Kroll, J. F. & Tokowicz, N. (2001). The development of conceptual representation for words in a second language. In J. L. Nicol (ed.), *One mind, two languages: Bilingual language processing*, pp. 49–71. Malden, MA: Blackwell.

Lasnik, H. (1999). *Minimalist analysis*. Malden, MA: Blackwell.

Laudanna, A. & Burani, C. (1985). Address mechanisms to decomposed lexical entries. *Linguistics*, 23, 775–792.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Lewis, R. L. (2000). Falsifying serial and parallel parsing models: Empirical conundrums and an overlooked paradigm. *Journal of Psycholinguistic Research*, 29, 241–248.

Lightfoot, D. (1991). *How to set parameters: Arguments from language change*. Cambridge, MA: MIT Press.

MacDonald, M. C., Pearlmutter, N. J. & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, 101, 676–703.

MacWhinney, B. (1987). The competition model. In B. MacWhinney (ed.), *Mechanisms of language acquisition*, pp. 249–308. Hillsdale, NJ: Erlbaum.

MacWhinney, B. (2001). The competition model: The input, the context, and the brain. In P. Robinson (ed.), *Cognition and second language instruction*, pp. 69–90. Cambridge: Cambridge University Press.

MacWhinney, B. & Bates, E. (eds.) (1989). *The crosslinguistic study of sentence processing*. Cambridge: Cambridge University Press.

Marslen-Wilson, W., Tyler, L. K., Waksler, R. & Older, L. (1994). Morphology and meaning in the English mental lexicon. *Psychological Review*, 101, 3–33.

McClelland, J. L. & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception, part 1: An account of basic findings. *Psychological Review*, 88, 375–407.

McClelland, J. L., St. John, M. & Taraban, R. (1989). Sentence comprehension: A parallel distributed processing approach. *Language and Cognitive Processes*, 4, SI 287–335.

Meng, M. & Bader, M. (2000). Ungrammaticality detection and garden path strength: Evidence for serial parsing. *Language and Cognitive Processes*, 15, 615–666.

Mitchell, D. C., Corley, M. M. B. & Garnham, A. (1992). Effects of context in human sentence parsing: Evidence against a discourse-based proposal mechanism. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 69–88.

Miyake, A. & Shah, P. (eds.) (1999). *Models of working memory: Mechanisms of active maintenance and executive control*. Cambridge: Cambridge University Press.

Morgan, J. L. & Demuth, K. (eds.) (1996). *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*. Mahwah, NJ: Erlbaum.

Nicol, J. L. & Pickering, M. J. (1993). Processing syntactically ambiguous sentences: Evidence from semantic priming. *Journal of Psycholinguistic Research*, 22, 207–237.

Nooteboom, S., Weerman, F. & Wijnen, F. (eds.) (2002). *Storage and computation in the language faculty*. Dordrecht: Kluwer.

Ouhalla, J. (1991). *Functional categories and parametric variation*. London: Routledge.

Pickering, M. J., Branigan, H. P., Cleland, A. A. & Stewart, A. J. (2000). Activation of syntactic information during language production. *Journal of Psycholinguistic Research*, 29, 205–216.

Pinker, S. (1984). *Language learnability and language development*. Cambridge, MA: Harvard University Press.

Pinker, S. (1987). The bootstrapping problem in language acquisition. In B. MacWhinney (ed.), *Mechanisms of language acquisition*, pp. 399–441. Hillsdale, NJ: Erlbaum.

Platzack, C. (1996). The initial hypothesis of syntax: A minimalist perspective on language acquisition and attrition. In H. Clahsen (ed.), *Generative perspectives on language acquisition*, pp. 369–414. Amsterdam: Benjamins.

Polinsky, M. (1997). American Russian: Language loss meets language acquisition. In W. Brown, E. Dornisch, N. Kondrashova & D. Zec (eds.), *Annual Workshop on Formal Approaches to Slavic Linguistics: The Cornell Meeting 1995*, pp. 370–406. Ann Arbor, MI: Slavic Publications.

Poulisse, N. (1999). *Slips of the tongue: Speech errors in first and second language production*. Amsterdam: Benjamins.

Poulisse, N. & Bongaerts, T. (1994). First language use in second language production. *Applied Linguistics*, 15, 36–57.

Rayner, K., Garrod, S. & Perfetti, C. A. (1992). Discourse influences during parsing are delayed. *Cognition*, 45, 109–139.

Salthouse, T. A. & Meinz, E. J. (1995). Ageing, inhibition, working memory, and speed. *Journals of Gerontology: Psychological Sciences*, 50B, 297–306.

Schmidt, R. & Frota, S. N. (1986). Developing basic conversational ability in a second language: A case study of an adult learner of Portuguese. In R. R. Day (ed.), *Talking to learn: Conversation in second language acquisition*, pp. 237–326. Rowley, MA: Newbury.

Schwartz, B. D. & Gubala-Ryzak, M. (1992). Learnability and grammar reorganization in L2A: Against negative evidence causing the unlearning of verb movement. *Second Language Research*, 8, 1–38.

Schwartz, B. D. & Sprouse, R. A. (1996). L2 cognitive states and the full transfer/full access model. *Second Language Research*, 12, 40–72.

Schwartz, R. G. & Terrell, B. Y. (1983). The role of input frequency in lexical acquisition. *Journal of Child Language*, 10, 57–64.

Sharwood Smith, M. (2001). The unsteady state: An SLA perspective on first language attrition. Paper presented at the Third International Symposium on Bilingualism (ISB3), Bristol, UK, April 2001.

Sharwood Smith, M. (in press). In two minds about grammar: On the interaction of linguistic and metalinguistic knowledge in performance. *Transactions of the Philological Society*, 102.

Slobin, D. I. (1985). Crosslinguistic evidence for the language-making capacity. In D. I. Slobin (ed.), *The crosslinguistic study of language* acquisition, vol. 2: *Theoretical issues*, pp. 1157–1256. Hillsdale, NJ: Erlbaum.

Steedman, M. & Altmann, G. (1989). Ambiguity in context. *Language and Cognitive Processes*, 4, SI 105–122.

Swinney, D. A. & Hakes, D. T. (1976). Effects of prior context upon lexical access during sentence comprehension. *Journal of Verbal Learning and Verbal Behavior*, 15, 681–689.

Toribio, J. A. (2000). On native language decline. Paper presented at the 25th Annual Boston University Conference on Language Development, November 2000.

Trahey, M. & White, L. (1993). Positive evidence preemption in the second language classroom. *Studies in Second Language Acquisition*, 15, 181–204.

Traxler, M. J. & Pickering, M. J. (1996). Plausibility and the processing of unbounded dependencies: An eye-tracking study. *Journal of Memory and Language*, 35, 454–475.

Truscott, J. & Wexler, K. (1989). Some problems in the parametric analysis of learnability. In R. J. Matthews & W. Demopoulos (eds.), *Learnability and linguistic theory*, pp. 155–176. Dordrecht: Kluwer.

Tyler, L. K. & Marslen-Wilson, W. D. (1977). The on-line effects of semantic context on syntactic processing. *Journal of Verbal Learning and Verbal Behavior*, 16, 683–692.

Uziel, S. (1993). Resetting Universal Grammar parameters: Evidence from second language acquisition of subjacency and the Empty Category Principle. *Second Language Research*, 9, 49–83.

Vainikka, A. M. & Young-Scholten, M. (1996a). The early stages in adult L2 syntax: Additional evidence from Romance speakers. *Second Language Research*, 12, 140–176.

Vainikka, A. M. & Young-Scholten, M. (1996b). Gradual development of L2 phrase structure. *Second Language Research*, 12, 7–39.

Valian, V. (1990). Logical and psychological constraints on the acquisition of syntax. In L. Frazier & J. de Villiers (eds.), *Language processing and language acquisition*, pp. 119–145. Dordrecht: Kluwer.

Weissenborn, J. & Höhle, B. (eds.) (2001). *Approaches to bootstrapping: Phonological, lexical, syntactic and neurophysiological aspects of early language acquisition* (vol. 1). Amsterdam: Benjamins.

White, L. (1989). *Universal Grammar and second language acquisition*. Philadelphia: Benjamins.

White, L. (1991). Adverb placement in second language acquisition: Some effects of positive and negative evidence in the classroom. *Second Language Research*, 7, 133–161.

White, L. (2001). Revisiting fossilization: The syntax/morphology interface. Paper presented to Edinburgh University Linguistic Circle, March 2001.

Zobl, H. (1990). Evidence for parameter-sensitive acquisition: A contribution to the domain-specific versus central processes debate. *Second Language Research*, 6, 39–59.

Received January 2, 2002
Revisions received January 2, 2003; September 17, 2003
Revision accepted October 24, 2003

PEER COMMENTARIES

Give syntax a chance

DEREK BICKERTON
Department of Linguistics, University of Hawaii, HI 96822, USA.
E-mail: bickertond@prodigy.net

As an occasional visitor to the land of SLA, I found myself somewhat mystified by the approach of Truscott and Sharwood Smith (henceforth TSS). Unless I have totally misunderstood them, they are arguing against the separate existence of both a Language Acquisition Device (LAD) and a Universal Grammar (UG). But who ever thought they were two? Occasionally linguists may write about one rather than the other, but I have always assumed them to be the same thing appearing in different guises (like Christ and the Holy Ghost in Christian theology). If you have UG, why would you need LAD? Processing with the aid of UG seems the only way to go – to the extent that UG is available to SLA, an issue TSS does not directly address.

TSS has a particular variety of processing in mind, and I am simply not familiar enough with competing models to be able to evaluate that variety. However, I feel uneasy with any approach that deals with issues in such abstract terms and contains no more than three example sentences (granted that this now seems *de rigeur* in Minimalist circles). My disquiet is increased when even these few examples are interpreted the way TSS interprets them.

TSS is far from being alone in underestimating the role syntax plays in determining semantic content (source of the widespread belief that a really good parser is – perhaps permanently – beyond the reach of science). But according to Hornstein (1999, p. 45) ‘semantic structure is a by-product of grammatical operations driven by formal concerns’, and I would endorse this strong statement as an excellent research strategy if falsifiability is a prime concern (if conventional beliefs turn out to be correct, counter-examples should pop up everywhere).

Take the first example, ‘It was only apparent, not real’, where TSS claims that ‘rejection of the wrong candidate [“a parent”, DB] only occurs at a late stage’ (p. 6). This statement is made *ex cathedra*, with no experimental evidence cited in support. In fact it seems to me inherently unlikely. ‘It was . . .’ should immediately cue the expectation of a non-human referent, since ‘it’ is neuter and the copula, if not immediately followed by a locative or adjectival, can only introduce a noun-phrase sharing the phi-features of the subject. There remains the possibility of sentences like those of (1) where ‘parent’ is immediately followed by either a participle or a complementizer:

(1) a. It was only a parent trying to locate her child.
b. It was only a parent that complained, not a teacher.

Sentences of either type would entail prior context – say, a phone call for (1a), previous discussion of some incident for (1b). Without this, and with comma intonation, readings with ‘a parent’ would be strongly disfavored and likely ruled out altogether, long before recourse to any conceptual processor.

We are then told that the syntactic processor might ‘try to construct a representation for “The horse is beautiful” in which *horse* is an adverb, like *really*, and the subject N is non-overt, along the lines of “(These creatures are ugly, but) those [$_N$e] really are beautiful”’ (p. 7).

While determiners such as demonstratives or numbers can license an empty category, unaccompanied articles cannot. That this is not an idiosyncrasy of English is shown by (2a), from Spanish, which roughly parallels TSS’s sentence:

(2) a. Estos animales son feos, pero los otros/los mios/*los son verdaderamente guapos
b. These creatures are ugly, but the others/mine/*the really are beautiful

Similar examples could be adduced from many other languages. Moreover, NP-deletion can only be licensed by a closely-preceding overt version of that NP, absent in the case of the ‘horse’ sentence. Since nouns are far more frequent in speech than adverbs, the default interpretation would in any case indicate a nominal; an adverbial reading would require positive evidence of adverbial identity.

To further illustrate how the syntactic processor works, take a sentence like (3), which actually appeared as a headline in a Denver newspaper, and which on the conventional view ought to boggle automatic translators, parsers and the like:

(3) Spy charges dog inspectors.

Here, three of the four words are ones that could in principle be either nouns or verbs. I assume that TSS’s syntactic processor would provide at least six possible readings. Mine gave me only two:

(4) a. A spy has made accusations against people who inspect dogs (or, has charged such people a fee).
b. Charges that they are spies have been persistently made against some inspectors.

(My pragmatic processor, primed by CNN, enabled me to select the second.) I think how my syntactic processor worked in this case was as follows.

(a) In a sentence of four words where no putative verb takes a sentential complement and there are no conjunctions or participial forms, there can be only one verb.
(b) The verb cannot be the first word because the remaining fragment, 'charges dog inspectors' is gibberish if all three words are nouns and the first is plural. The latter fact, taken in conjunction with the lexical items chosen here, sharply reduces the chances that 'charges' could be being used attributively, as 'tariffs' is in 'tariffs debate closure'.
(c) Agreement is okay if the second word is a verb ('spy' sing., 'charges' sing.) and if the third word is a verb ('charges' pl., 'dog' pl.) yielding the two remaining possibilities, (4a) and (4b), as possible readings.

In other words, even if Hornstein's dictum proves overly strong, the role of the syntactic processor in sentence processing is hard to overestimate, and TSS, like most work, errs in the opposite direction. For all that, TSS's version of a processing model may still beat the competition. I wouldn't know; I'm a stranger here myself.

Reference

Hornstein, N. (1999). Minimalism and quantifier raising. In S. D. Epstein & N. Hornstein (eds.), *Working minimalism*, pp. 45–75. Cambridge, MA: MIT Press.

Acquisition by Processing Theory: A theory of everything?

SUSANNE E. CARROLL
Institute for English & American Studies, Universität Potsdam, PF 60 15 53, D-14415 Potsdam, Germany
E-mail: carroll@rz.uni-potsdam.de

Truscott and Sharwood Smith (henceforth T&SS) propose a novel theory of language acquisition, ACQUISITION BY PROCESSING THEORY (APT), designed to account for both first and second language acquisition, monolingual and bilingual speech perception and parsing, and speech production. This is a tall order. Like any theoretically ambitious enterprise, the APT shares certain properties with much that has gone before. Like the Competition Model (CM; MacWhinney, 1987, 1997; MacWhinney and Bates, 1989, inter alia) and other associative network connectionist learning models, the APT eschews a Language Acquisition Device (LAD) by treating acquisition as the strengthening of levels of representation activation. A parser can produce multiple representations of a parse string in parallel, which then 'compete' as analyses for an input string. Unlike the CM, however, the APT is not motivated by a solid program of empirical studies in language acquisition or cross-language processing. Nor does it strike me as theoretically coherent, for the APT, unlike the CM, assumes that knowledge of language involves knowledge of grammatical structure and that the parser makes deterministic use of Universal Grammar in the form of a Minimalist grammar. The determinism is important here; the claim to eliminate LAD hinges on it.

Like the Autonomous Induction Theory (AIT; Carroll, 1999, 2001, 2002a, b), the APT embeds acquisition in the tri-partite processing architecture of Jackendoff's (1987, 1990, 1997, 2002). What is not apparent from the APT, is that Jackendoff's research is not designed to motivate a particular theory of PARSING but, rather, to motivate the Representational Modularity Theory of linguistic competence, which explicitly rejects the syntacticocentric views of grammar which Chomsky has espoused. The empirical motivation for the tri-partite theory of competence undermines Minimalism. The APT would appear to require both PF and LF in order to explain a number of grammatical phenomena; these additional levels of syntactic representation are fundamentally incompatible with the Jackendovian model of linguistic competence.

The AIT, like the Modularity Matching Model (MMM; Crain and Thornton, 1998; Crain and Wexler, 1999), assumes that a LAD is responsible for creating the structural differences in grammars which, in turn, cause the well-documented behavioural differences in users of different languages. The AIT assumes that the LAD is severely constrained in numerous ways, but is essential to explaining the equally well-documented differences between child and adult grammars, or between L2 learners' and monolingual native speakers' grammars of the target language. It postulates a LAD for the simple reason that these differences do not reduce to differences in the specification, and combination, of lexical items. The APT cannot describe, let alone explain, well-documented language-specific NON-LEXICAL parsing preferences (Cuetos and Mitchell, 1988; Mitchell, Cuetos, Corley and Brysbaert, 1995; Cuetos, Mitchell and Corley, 1996), which turn out to be relevant for the parsing of L2 learners (Fernández, 2003). The APT cannot describe, let alone explain, processing differences arising between early and late L2 learners attributable to differences in the way in which each group of learners encodes a universal feature in the lexical entry of nouns (Guillemon and Grosjean, 2001).

Along with the CM, the AIT, the MMM, and emergentist approaches to syntax (O'Grady, 2002, 2003), the APT hypothesises that children and adults share a common language-processing system. Like emergentist models, the APT sees structural knowledge emerging automatically through the operation of language processors building representations, using information from lexical items to do so. In the case of the APT, it is Merge. Like the MMM, the APT assumes that language acquirers have access to a universal, invariant parser and that all learners have access to the constraints and principles of UG at all stages of acquisition. Crain and Thornton (1998) make use of the Principles and Parameters setting version of generative syntax; the APT adopts Minimalism. The MMM, however, hypothesises that all of the linguistic abilities of the child are the same as the adults (Crain and Thornton, 1998, p. 30), including working memory and attention, and, therefore, that differences in learners' grammars cannot be influenced by such things. Thus, whenever the behaviour of children and adults differs on some task, the differences must be due to their grammars – and hence to the LAD. T&SS have no LAD; they do invoke some undocumented limits in working memory in 'older learners' (p. 15). One can only guess what population T&SS are referring to, but empirical research dealing with the effects of aging on working memory (Krampe, Engbert and Kliegl, 2001; Oberauer, Demmrich, Mayr and Kliegl, 2001;

Oberauer and Kliegl, 2001) suggests that differences in working memory are not going to support the differences in children and adult grammars in second language acquisition.

The APT does not arise from a response to specific empirical issues in either first or second language acquisition. The major goal appears to be a sweeping application of Occam's razor. But the savings are only apparent. Standard parsers implement acquired grammatical distinctions. T&SS introduce several representation-creating and representation-changing operations into their parser: it creates 'empty' SS and CS or PS items by coindexation, writes features onto lexical representations, deletes features from lexical representations ('reversing' prior feature assignment), copies 'chunks' of representations. None of these operations are deterministic, formally well-defined, nor do they appear to be constrained.

Minimalism without empirical and descriptive adequacy is hardly an advantage. Consider what T&SS give up.

(i) By attempting to reduce parsing to the activation of lexical items, they lose an account of cross-linguistic parsing differences.

(ii) With Merge and activation as its procedures, the APT cannot describe, let alone explain, the acquisition of numerous language-specific constructions which are not reducible to c- or s-selection properties of lexical items, including Heavy-NP Shift, Left and Right Dislocations, parenthicals, and many more (Jackendoff, 2002, pp. 178–183).

(iii) By apparently confounding primitive linguistic features with lexical items, and by failing to distinguish language parsing from language production, T&SS cannot describe, let alone explain, how bilingual learners create and use unique word classes, particularly classes only indirectly related to the speech they hear. Deuchar (1999), for example, argues that her learner treats *más paper* 'more paper' or *oh-dear book* 'an unexpected event has affected the book' as a predicate (argument) structure, the forms *más* and *oh-dear* being acategorial, i.e., non-lexical categories. Deuchar explicitly states that expressions like *más paper* are not part of the adult input so we must conclude that the child's parser is never confronted with it. These utterances arise not because of the way the child's parser parses *más paper* but rather as a result of the child's combining bits of language in an entirely creative way. The model appears to be motivated by universal primitives available in the logical structure of Conceptual Representations, namely functors and arguments. Such creativity cannot be the sole result of 'lexical activation' since that account fails to explain how the category arises in the first place. Note too that these word classes drop out of the child's grammar in their original predicate functions, or disappear entirely. The predicate *oh-dear* thus is reclassified into an interjection (with no combinatorial properties). Since the form *oh-dear* survives into the adult system, it must continue to be activated. But the formative loses some conceptual features and acquires others over the course of time and loses all of its syntactic features. The APT cannot capture such developmental patterns.

(iv) By failing to recognise the fundamental differences between parsing and speech production, T&SS have serious problems explaining where grammatical competence ends and performance begins. The sentence grammars of English and French both require a characterisation of left dislocations (Carroll, 1981; Lambrecht, 1981, 1994), but a theory of speech production must explain how utterances like the following arise:

(1) a. Moi-là, c'est Poulet Doré
Me-interjection it is Poulet Doré [the name of a restaurant].
'I prefer Poulet Doré' (Carroll, 1981)

b. On a dû se battre pour la France aussi.
one PAST ought oneself fight for the France too
Ça on a jamais refusé, quoi mais euh moi,
that one PAST never refused, what but um me,
je tiens compte de la Bretagne
I hold account of the Brittany
'we had to fight for France too, that we never refused, you know, but me, I take Brittany into consideration' (cited in Beeching, 1999, p. 82)

c. Well, that guy, you know, the one who's holding up the bar, he comes here every night.

The APT is certainly innovative, but its innovation – namely the elimination of the LAD and the apparent reduction of processing to changes of activation levels – is neither warranted by acquisition data nor appears to be an improvement over other theories. In its current scope, the APT looks like a theory of everything, and unlikely to meet standard criteria of adequacy of a theory of language acquisition.

References

Beeching, K. (1999). Update on spoken French. *Bilingualism: Language & Cognition*, 2(2), 81–83.

Carroll, S. E. (1981). *Notions fonctionnelles en grammaire transformationnelle: Dislocations et structures topicalisées en français contemporain*. PhD dissertation, Université de Montréal.

Carroll, S. E. (1999). Putting 'input' in its proper place. *Second Language Research*, 15(4), 337–388.

Carroll, S. E. (2001). *Input and evidence: The raw material of second language acquisition*. Amsterdam: Benjamins.

Carroll, S. E. (2002a). I-learning. In S. Foster-Cohen, T. Ruthenberg & M.-L. Poschen (eds.), *EUROSLA Yearbook 2*, pp. 7–28. Amsterdam: Benjamins.

Carroll, S. E. (2002b). Induction in a modular learner. *Second Language Research*, 18(3), 224–249.

Crain, S. & Thornton, R. (1998). *Investigations in Universal Grammar: A guide to experiments on the acquisition of syntax and semantics*. Cambridge, MA: MIT Press.

Crain, S. & Wexler, K. (1999). Methodology in the study of language acquisition: A modular approach. In W. Ritchie & T. Bhatia (eds.), *Handbook of child language acquisition*, pp. 387–425. San Diego, CA: Academic Press.

Cuetos, F. & Mitchell, D. C. (1988). Cross-linguistic differences in parsing: Restrictions on the use of the Late Closure Strategy in Spanish. *Cognition*, 30, 73–105.

Cuetos, F., Mitchell, D. C. & Corley, M. M. B. (1996). Parsing in different languages. In M. Carreiras, J. E. Garcia-Albea & N. Sebastián-Gallés (eds.), *Language processing in Spanish*, pp. 145–90. Hillsdale, NJ: Erlbaum.

Deuchar, M. (1999). Are function words non-language-specific in early bilingual two-word utterances? *Bilingualism: Language & Cognition*, 2(1), 23–34.

Fernández, E. (2003). *Bilingual sentence processing: Relative clause attachment in English and Spanish*. Amsterdam: Benjamins.

Guillemon, D. & Grosjean, F. (2001). The gender-marking effect in spoken word recognition: The case of bilinguals. *Memory & Cognition*, 29(3), 503–511.

Jackendoff, R. (1987). *Consciousness and the computational mind*. Cambridge, MA: MIT Press.

Jackendoff, R. (1990). *Semantic structures*. Cambridge, MA: MIT Press.

Jackendoff, R. (1997). *The architecture of the language faculty*. Cambridge, MA: MIT Press.

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

Krampe, R. Th., Engbert, R. & Kliegl, R. (2001). Age-specific problems in rhythmic timing. *Psychology & Aging*, 16, 12–30.

Lambrecht, K. (1981). *Topic, antitopic and verb agreement in non-standard French*. Amsterdam: Benjamins.

Lambrecht, K. (1994). *Information structure and sentence form*. Cambridge: Cambridge University Press.

MacWhinney, B. (1987). Applying the Competition Model to bilingualism. *Applied Psycholinguistics*, 8, 315–327.

MacWhinney, B. (1997). Second language acquisition and the Competition Model. In A. M. B. de Groot & J. F. Kroll (eds.), *Tutorials in bilingualism: Psycholinguistic perspectives*, pp. 113–142. Mahwah, NJ: Erlbaum.

MacWhinney, B. & Bates, E. (eds.) (1989). *The cross-linguistic study of sentence processing*. Cambridge: Cambridge University Press.

Mitchell, D. C., Cuetos, F., Corley, M. M. B. & Brysbaert, M. (1995). Exposure-based models of human-parsing: Evidence for the use of course-grained (non-lexical) statistical records. *Journal of Psycholinguistic Research*, 24(6), 469–488.

Oberauer, K., Demmrich, A., Mayr, U. & Kliegl, R. (2001). Dissociating retention and access in working memory: An age-comparative study of mental arithmetic. *Memory & Cognition*, 29, 18–33.

Oberauer, K. & Kliegl, R. (2001). Beyond resources – Formal models of complexity effects in age differences in working memory. *European Journal of Cognitive Psychology*, 13, 187–215.

O'Grady, W. (2002). An emergentist approach to syntax. Ms., University of Hawai'i.

O'Grady, W. (2003). Language without grammar. Plenary to the Annual Meeting of the American Association of Applied Linguistics, March 23, 2003, Arlington, VA.

Bilingualism: Language and Cognition 7 (1), 2004, 26–28 © 2004 Cambridge University Press DOI: 10.1017/S1366728904001208

APT: Costs and benefits of a hybrid model

TON DIJKSTRA
Nijmegen Institute for Cognition and Information (NICI), University of Nijmegen, P.O. Box 9104, 6500 HE Nijmegen, The Netherlands
E-mail: dijkstra@nici.kun.nl
MARCO HAVERKORT
Department of Linguistics, University of Nijmegen
Boston University

In their keynote contribution, Truscott and Sharwood Smith offer a general model of language development from a processing perspective. As they state, their model is very ambitious: Their 'acquisition by processing' theory (APT) aims not only at explaining both first and second language acquisition but also real-time processing in language comprehension and production. APT takes a cross-disciplinary approach that intends to bring together research on linguistic structure and on general cognition. However, the joint contribution of linguistic and psycholinguistic approaches is mainly evident in terms of theoretical concepts (e.g. UG, syntactic rules, modules, activation) because the presented empirical evidence is limited in scope.

The authors' attempt at developing an interdisciplinary approach is laudable because it may lead to a model that benefits from insights and strengths of several disciplines. However, such an approach may be vulnerable because of limitations of each contributing approach and it may suffer from incompatibilities between paradigmatic principles. In the following sections, we will consider the extent to which APT appears to profit from cross-disciplinary integration while avoiding its drawbacks.

Before presenting our comments, we note that our understanding of APT may be less than perfect because the current paper tells a rather global and abstract tale with relatively few concrete examples and applications. In addition, unclarities sometimes arise because of loose wording. We consider it to be of the utmost importance that the terminology used in formulating a new theory is as unambiguous and explicit as possible. However, in the section 'The place of an L2 in the language module', the authors state that 'two knowledge bases are competing for access to a single processing system' (p. 4). We take this to mean that a presented input leads to competition of some kind between information originating from two knowledge bases. Knowledge bases, being passive collections of representations, cannot really compete themselves; only activated representations can compete. Is there even independent evidence for two separate knowledge bases? Clearly, one integrated knowledge base cannot compete with itself. As a second example, note that the authors maintain that in the example 'They walk to work' (one of the few they give in the paper), there is 'nothing in the phonological form of the utterance [that] indicates its [INFL's] presence' (p. 5). The subject pronoun, however, appears in the nominative form, indicative of the presence of INFL, which assigns nominative case when it is finite or, in minimalist terms, checks nominative case.

Although imprecise wording may lead to misunderstandings, we are more concerned with problems arising from tension between the paradigmatic principles from linguistic or psycholinguistic approaches. At the risk of oversimplifying, APT appears to be a linguistic theory on which a psychological learning mechanism and psycholinguistic processing notions have been grafted. Most of our points of criticism follow from the tension that arises from the specific way in which the different approaches are combined, and omissions in the operation.

1. Separate knowledge representations and processing mechanisms

In APT, a distinction is made between knowledge representations and processing mechanisms, at least with respect to lexical aspects of linguistic knowledge. Such a distinction is indeed important from a psycholinguistic point of view, and potentially allows an extension of the authors' account of second language learning to performance by language user populations that exhibit deviant language use. The spontaneous speech of aphasics, for instance, shares a number of characteristics with that of young children and second language learners, and would therefore be a prime candidate for such a treatment.

Indeed, there is evidence from aphasia that a strict distinction should be made between the representation of linguistic knowledge and the use of that knowledge in language processing. Aphasics appear to retain the grammatical knowledge of their language, but limited working memory capacity and timing problems prevent a normal use of this knowledge in language production and comprehension (cf. Kolk, 1995; Sabourin and Haverkort, 2003). Evidence for a similar distinction has recently been presented for second language learners (Sabourin, 2003).

However, it is unclear whether the APT account can be applied to all other (e.g. non-lexical) types of linguistic knowledge. It has been shown, for instance, that aphasics still exhibit semantic priming effects at trace positions of moved elements, be it with a delay of about 600 ms, due to a delay in processing (Burkhardt, Piñango and Wong, 2003). This indicates that they establish filler-gap relations. Arguably, these relations are established in the universal syntax processor of the model, so it would have to be assumed that this component and the lexical representations (which can be shown to be available to aphasics by means of syntactic priming effects, cf. Haarmann, 1992) can be available at the same time; in that case, processing problems must be due to working memory constraints.

2. The role of working memory constraints

Unfortunately, working memory is the least worked-out aspect of the model. Its structure and precise role are left pretty much implicit. This is regrettable, especially because it is precisely this component that is so central to explaining deviant language behavior in aphasics and, arguably, first and second language learners (see below).

Working memory is also a key feature of the model to account for the old competence–performance distinction. One needs it to explain why longer and syntactically more complex sentences cause more problems in production and comprehension. It may also partly explain why models of word recognition assume parallel activation of different lexical candidates, while those of syntactic parsing assume bottom-up and depth-first processing with only few syntactic structures active in parallel (cf. garden path sentences). As it is so central a notion from a number of different perspectives, its structure and role need to be made more explicit.

Here we illustrate just one aspect of its importance. Working memory differences may underlie individual differences in task execution. Indeed, in some cases performance differences between monolinguals and bilinguals may have nothing to do with differences in syntactic knowledge but with differences in working memory capacity. A pilot experiment conducted in the NICI lab illustrates this point (Caelen, 1998). When Dutch-German bilinguals processed German (L2) sentences with temporarily ambiguous subject- and object-relative clauses, they performed differently from either Dutch or German participants in their L1, who performed similarly. Both quantitative (latencies) and qualitative (interactions between syntax and semantics) differences in performance occurred. In contrast to monolinguals, Dutch-German bilinguals processing their L2 used semantic strategies to help them resolve syntactic ambiguities. It appears that this difference in processing strategies was a consequence of a relatively heavier load on the bilinguals' working memory during L2 sentence processing. In sum, task- and strategy-dependent mechanisms for normal processing are needed and must be specified.

Also from a DEVELOPMENTAL perspective, it is too bad that the role of working memory is not made more explicit and that the authors do not relate their notion of working memory to the models available in the literature (in this respect, it is telling that the seminal work of Alan Baddeley is missing from the bibliography: Baddeley, 1986, 2000; Gathercole and Baddeley, 1993). In APT, aspects of development can either be explained as lexical learning or as changes in working memory capacity. This follows from UG no longer being a set of universal principles, but being equated with the syntactic processor, plus constraints on the structure and types of lexical information. Assuming that the universal syntactic processor is innate, language learning is then restricted to learning lexical characteristics.

3. Activation in learning versus activation in processing

Tension arises because the authors combine all-or-none linguistic notions with respect to syntactic rules and feature representations with continuous psycholinguistic notions such as activation. With respect to the acquisition of syntactic and other rules, the authors provide some illustrations of their learning mechanism in the section 'Acquisition by processing: Functional categories'. One key mechanism is that repeated input may lead to a (gradual?) formulation of rules, using innate knowledge that is already available from UG in combination with an activation mechanism. Although their account is not specified in this respect, the authors appear to suggest that, by application of UG, certain discovered regularities ultimately lead to the instantiation of rules with exceptions. However, the use of UG to make sure that certain syntactic rules and not others are learned appears to entail an important constraint on the learning mechanism that is difficult to evaluate.

This issue reminds us of the morphological literature with respect to the acquisition of the past tense of strong verbs. The authors and Parallel Distributed Processing (PDP) connectionists superficially appear to share the theoretical view that there is only one learning mechanism for rule-following and exceptional cases. The authors, however, appear to hold that although there is only one principle for learning, not all possible rules are treated equally by UG. Thus, the question is: In what way is the activation level of regular and irregular cases determined by their token frequency (learning mechanism) and the type frequency of the rules they follow (UG)? In the model, activation strength does not appear to be just a function of frequency of usage, as a learning mechanism would seem to suggest (and as connectionists would argue).

A similar tension arises with the introduction of the [strong] and [weak] features that are all-or-none notions but can vary in activation level. In Chomsky's Minimalist Program (Chomsky, 1995), features are privative in nature: they are either weak or strong, allowing checking to be postponed until after spell-out, or forcing it to take place before spell-out, respectively. However, in the present theory, features also vary in activation level. In our view, the privative weak/strong distinction is incompatible with the notion of activation level that the authors introduce. As we said above, the notion of activation level fits in better with a connectionist/PDP type model than with a rule-based model, which the authors are assuming.

4. Still other disciplines to be integrated?

In spite of APT's cross-disciplinary approach, important sources of evidence and theoretical notions appear to have been neglected. For instance, there is considerable recent evidence from the field of Cognitive Neuroscience, such as ERP measurements, that may be informative with respect to bilingual syntactic processing (Sabourin, 2003; for an overview, see Kroll and Dussias, 2003).

Furthermore, at the end of the paper, the authors remark that 'this process of change is inherent in the nature of activation levels . . . In adopting this view we diverge from all explicit theories we are familiar with.' (p. 17) Nevertheless, the learning mechanism proposed by APT has some striking commonalities with the principle of self-organization in Dynamic Systems Theory (DST), found in the work by, for instance, Edelman (Edelman, 1992; Reeke and Edelman, 1984), and Thelen and Smith (1994). In fact, we believe that important ideas about how this type of learning mechanism may operate can be found in available DST studies. For instance, in DST we find notions like the emergence of self-organization in non-supervised systems, context effects on category formation, and the development of knowledge in action. However, proponents of a DST approach would argue that the learning mechanism proposed by Truscott and Sharwoord Smith (using only context-specific adaptations) would only be optimal if a simplistic notion of 'innate ideas' is abandoned (cf. the work by Johnson and Newport, 1989; Newport, 1990, on bilingual learning).

To conclude, we argue that the proposed theory does not merely entail acquisition by processing, because of the constraints imposed on the acquisition process by UG. Furthermore, we propose that other constraining factors should be considered, for instance, working memory limitations and task-dependent strategies. These notions are of the utmost importance because they bridge the gap between competence and actual performance. Without their specification and empirical evidence, APT is a blueprint of a building for which the groundwork must still be tested.

References

Baddeley, A. D. (1986). *Working memory*. Oxford: Clarendon Press.

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, 4, 417–423.

Burkhardt, P., Piñango, M. & Wong, K. (2003). The role of the anterior left hemisphere in real-time sentence comprehension: Evidence from split intransitivity. Ms., Yale University.

Caelen, M. (1998). *Extending the study on the processing of relative clauses to bilingualism*. M.A. thesis, University of Nijmegen.

Chomsky, N. (1995). *The minimalist program*. Cambridge, MA: MIT Press.

Edelman, G. M. (1992). *Bright air, brilliant fire: On the matter of mind*. New York: Basic Books.

Gathercole, S. E. & Baddeley, A. D. (1993). *Working memory and language*. Hove: Lawrence Erlbaum.

Haarmann, H. (1992). *Agrammatic aphasia as a timing deficit*. PhD dissertation, University of Nijmegen.

Johnson, J. & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, 21, 60–99.

Kolk, H. (1995). A time-based approach to agrammatic production. *Brain and Language*, 50, 250–260.

Kroll, J. F. & Dussias, P. E. (2003). The comprehension of words and sentences in two languages. In T. Bhatia & W. Ritchie (eds.), *Handbook of bilingualism*, pp. 169–200. Cambridge, MA: Blackwell Publishers.

Newport, E. L. (1990). Maturational constraints on language learning. *Cognitive Science*, 14, 11–28.

Reeke, G. N. , Jr. & Edelman, G. M. (1984). Selective networks and recognition automata. *Annals of the New York Academy of Science*, 426, 181–201.

Sabourin, L. (2003). *Grammatical gender and second language processing: An ERP study*. PhD dissertation, University of Groningen.

Sabourin, L. & Haverkort, M. (2003). Neural substrates of representation and processing of a second language. In R. van Hout, A. Hulk, F. Kuiken & R. Towell (eds.), *The lexicon-syntax interface in second language acquisition*, pp. 175–197. Amsterdam & Philadelphia: John Benjamins.

Thelen, E. & Smith, L. B. (1994). *A dynamic systems approach to the development of cognition and action*. Cambridge MA: MIT Press.

Between the input and the acquisition lies the shadow

MICHAEL HARRINGTON
Linguistics, School of English, The University of Queensland, Michie Bldg Level 4, Brisbane 4072, Queensland, Australia
E-mail: mwharr@uq.edu.au

Acquisition by Processing Theory (APT) is a unified account of language processing and learning that encompasses both L1 and L2 acquisition. Bold in aim and broad in scope, the proposal offers parsimony and comprehensiveness, both highly desirable in a theory of language acquisition. However, the sweep of the proposal is accompanied by an economy of description that makes it difficult to evaluate the validity of key learning claims, or even how literally they are to be interpreted. Two in particular deserve comment; the first concerns the learning mechanisms responsible for adding new L2 grammatical information, and the second the theoretical and empirical status of the activation concept used in the model.

Learning

The APT adopts the modular processing architecture developed by Jackendoff (1997, 2002). Linguistic ability resides in a set of encapsulated but highly interactive (sub-) modules, e.g. the phonological store (PS), the syntactic store (SS) and conceptual store (CS) as well as interface processors (p. 2). Processing in both production and comprehension is characterized as the interaction of the processors that serves to modulate the activation level of the representations involved, the resulting levels determining what representations are used in processing the message and in learning. Two types of learning are posited. One is the strengthening of the resting activation level of representations already present, making them more available for future use. Frequency effects are the most obvious manifestation of this kind of learning. The other involves the learning of new L2 grammatical knowledge and the restructuring of the L2 grammar that results. Learning is said to occur when the various processors cannot construct 'an adequate representation' of the current input and are thus forced to add a new item (p. 6). What an item consists of is not specified, though it includes such sub-lexical elements as functional categories, e.g. inflection (I). Here the learner would need to associate a form encountered in the input (e.g. the copula *be*) with the syntactic representation. But as there are no direct surface cues for I, the mapping between the surface item and the underlying syntax must involve a relatively abstract level of representation. This arises in the APT as surface phonological and conceptual information is extracted and the respective processors 'coindex' the various mappings with the output of the SS to yield an acceptable representation. How these mappings occur, and how the various sub-processors and the system as a whole assesses the adequacy of the resulting representations, is not clear.

It is here that superficial, but misleading, similarities can be drawn between the APT and the Competition Model (CM; Bates and MacWhinney, 1989; MacWhinney, 1997). Linguistic knowledge in the CM is characterized as a highly complex but nonetheless direct set of mappings between surface forms (cues) and underlying functions in a specific language. The CM explicitly rejects, a priori, an autonomous level of syntactic representation (Bates and MacWhinney, 1989, p. 39). The importance of the mappings is determined by their distribution in the language, which in turn represents the probability with which they will be learned. In the paradigmatic Competition Model task the interaction of various cue strengths (preverbal position, animacy, contrastive stress, etc.) predict cross-linguistic judgments of agency. Simple associative learning mechanisms are responsible for developing these cue strengths, which define language specific processing strategies (Harrington, 1987). They ARE the formal characteristics of the language in the CM framework and are not independent elements that must be acquired separately (p. 9). Like the APT, the CM assumes that learning is the direct result of processing, unlike the latter, the specific mechanism responsible for making that form–function mapping is not identified in the APT. How does one develop probabilistic cue strengths for innately specified principles, as in the case of functional categories?

To be sure, the APT escapes the considerable induction problem facing input-driven approaches by assuming UG as a direct constraint on learning (Harrington and Dennis, 2002). But the a priori assumption of a guiding role in learning for these abstract linguistic principles is at fundamental odds with the aims and the logic of the CM approach. Furthermore, the formal theory APT adopts to constrain this process, a mixture of Principles & Parameters and Minimalism, makes no claims in regards to either processing or learning: 'The ordering of operations [in grammatical theory] is abstract, expressing postulated properties of the language faculty of the brain, with no temporal interpretation implied.' (Chomsky, 1995, p. 308). That no interpretation is implied doesn't mean none exists; however, it will be necessary for the

APT to provide an account of how this abstract knowledge, assumed to be directly embodied in the processor, will constrain the extraction of knowledge from the input in language performance. This remains the central issue in second language acquisition (Carroll, 2001).

Activation

The concept of activation is central to the APT. The level of activation of specific representations is the dependent variable in learning in the account, but how literally we are meant to take the notion is unclear. Activation can be used as a descriptive label to characterize learning outcomes, as in the raising or lowering of an item's activation level as a result of exposure (p. 5). The use of activation in this sense has no explanatory value. On the other hand, if a neurolinguistic interpretation is implied, then the unity of L1 and L2 processes stipulated in the APT is questionable. In a widely-cited study examining language representation in the bilingual brain, Kim, Relkin, Lee and Hirsch (1997) found evidence for distinctly different areas of cortical activation in language use in early and late bilinguals. Functional MRI data from sentence generation tasks showed that L2 activation for adult bilinguals, who learned the L2 post-critical period, occurred in an anatomically separate area of the brain. As the task used in the study would certainly have involved activation of the kind of representations central to the APT account, the separation is a problem. Although research in this area has yielded conflicting findings (e.g. Illes, Francis, Desmond, Gabrieli, Glover, Poldrack, Lee and Wagner, 1999), the evidence that exists for the separation of the L1 and L2 argues against a unified model of language at the cortical level.

The empirical evidence for the activation construct is from the structural priming literature (Bock and Loebell, 1990, for an overview see Pickering and Branigan, 1999). This research has demonstrated significant effects of prior exposure to subsequent production of structures, independent of lexical and semantic effects. Although interesting, significant questions remain concerning the occurrence and persistence of these effects, the nature of the structural representations being activated, and the mechanisms responsible for the priming. Priming in experimental studies, while occurring significantly above chance, is far from systematic. Pickering and Branigan (1999) report priming in confederate–participant conversations to range for from 78% for the same-verb primes (chance is 50%) and 63% for different verbs. Also, there is mixed evidence as to how long such activations persist. Branigan, Pickering and Cleland (1999) report very short-lived activation, while most studies show the priming effect to persist for some time (Chang, Dell, Bock and Griffin, 2000; up to one week in the case of aphasic patients in Saffran and Martin, 1997). As Chang et al. (2000, p. 219) note, the persistence of priming over time and intervening processing argues against a short-term memory or temporary activation effect, and rather reflects longer-term effects in memory.

A brief activation period is crucial for the APT account. It is assumed that during processing the encapsulated SS will automatically generate two (or more) possible analyses for the input it receives from the other processors. The unacceptable parses must be removed immediately and the correct alternative returned as output activation. This process is particularly important for the addition of new items whose resting level must attain a certain value to be used (pp. 5–6). To test this claim the authors propose a hypothetical experiment. In the study a subject would be primed with a sentence like 'The floor was uncovered wood'. Two candidate parses are available here for subsequent priming. The correct one interprets 'was' as the main verb and the other interprets it as part of the passive, prompted by the temporary ambiguity between the verb and adjectival reading of 'uncovered'. Both will be momentarily activated, with the alternative passive quickly rejected. The authors argue that evidence for this crucial removal mechanism would be the absence of subsequent priming by the rejected analysis. In other words, prior exposure to 'The floor was uncovered wood' would not result in the subsequent production of passive structures. Given the null-effects logic of the proposed study and the fact that we have to accept the more central (though quite defensible) assumption concerning the activation of competing structures, one might question how persuasive such findings would be as support for the account and the theoretical framework on which it rests.

Which is just as well for the APT, given that robust evidence for priming by the rejected parse already exists. Bock and Loebell (1990) showed that sentences containing an adjunct 'by' phrase (e.g. 'The foreigner was loitering by the traffic lights') often primed the production of the passive alternative (e.g. 'The boy was stung by the bee').

These findings also underscore another problem with the structural priming evidence, and this concerns the nature of the knowledge representations being activated. The Bock and Loebell (1990) results show that the effect is structural, but that the structures activated involve surface level configurations. Structures involving argument phrase structures ('by the bee') can prime adjunct phrase structures ('by the traffic lights'). How these surface effects reflect the working of the (innately-specified) competence grammar needs to be specified.

Structural priming has been observed and studied in the context of the language production system and the various demands, linguistic and interactional, that production places on the speaker. Although recent work has sought to establish priming as the result of the nature of language architecture shared by comprehension and production processes (Pickering, Branigan, Clelan and Stewart, 2000), and possibly even implicated in implicit learning (Chang et al., 2000), it remains to be seen whether the effect is due to production processes per se or arises as a more fundamental feature of the language processing system.

A unified account of learning and processing

The authors lament what they see as the increasing fragmentation in SLA research and theory that seems to make a unified theory increasingly difficult, if not impossible. This problem is not unique to SLA, as specialization has made conceptual syntheses increasingly difficult across all scientific disciplines (Wilson, 1998). The APT is intended as a cross-disciplinary platform that illuminates the key issues by bringing together various strands of research relevant to SLA. This is a laudable aim but one that appears unattainable in the near term, given the complexity of the various domains – linguistic, psycholinguistic, and cognitive – that must inform such a model.

References

Bates, E. & MacWhinney, B. (1989). Functionalism and the Competition Model. In B. MacWhinney & E. Bates (eds.), *The cross-linguistic study of sentence processing*, pp. 3–73. Cambridge: Cambridge University Press.

Bock, J. K. & Loebell, H. (1990). Framing sentences. *Cognition*, 35, 1–39.

Branigan, H. P., Pickering, M. J. & Cleland, A. A. (1999). Syntactic priming in written production: Evidence for rapid decay. *Psychonomic Bulletin and Review*, 6, 635–640.

Carroll, S. E. (2001). *Input and evidence: The raw material of second language acquisition*. Amsterdam: John Benjamins.

Chang, F., Dell, G. S., Bock, K. & Griffin, Z. M. (2000). Structural priming as implicit learning: A comparison of models of sentence production. *Journal of Psycholinguistic Research*, 29, 217–229.

Chomsky, N. (1995). *The Minimalist program*. Cambridge, MA: MIT Press.

Harrington, M. (1987). Processing transfer: Language specific processing strategies as a source of interlanguage variation. *Applied Psycholinguistics*, 8, 351–377.

Harrington, M. & Dennis, S. J. (2002). Input-driven language learning. *Studies in Second Language Acquisition*, 24, 261–268.

Illes, J., Francis, W. S., Desmond, J. E., Gabrielli, J. D. E., Glover, G. H., Poldrack, R., Lee, C. J. & Wagner, A. D. (1999). Convergent cortical representation of semantic processing in bilinguals. *Brain and Language*, 70, 347–363.

Kim, K. H. S., Relkin, N. R., Lee, K. M. & Hirsch, J. (1997). Distinct cortical areas associated with native and second languages. *Nature*, 388, 171–174.

Jackendoff, R. (1997). *The architecture of the language faculty*. Cambridge, MA: MIT Press.

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

MacWhinney, B. (1997). Second language acquisition and the Competition Model. In A. M. B. de Groot & J. F. Kroll (eds.), *Tutorials in bilingualism: Psycholinguistic perspectives*, pp. 113–142. Mahwah, N. J.: Erlbaum.

Pickering, M. J. & Branigan, H. P. (1999). Syntactic priming in production. *Trends in cognitive science*, 3, 136–141.

Pickering, M. J., Branigan, H. P., Clelan, A. A. & Stewart, A. J. (2000). Activation of syntactic information during language production. *Journal of Psycholinguistic Research*, 29, 205–216.

Saffran, E. M. & Martin, N. (1997). Effects of structural priming on sentence production in aphasics. *Language and Cognitive Processes*, 12, 877–882.

Wilson, E. O. (1998). *Consilience: The unity of knowledge*. New York: Knopf.

Bilingualism: Language and Cognition 7 (1), 2004, 32–34 © 2004 Cambridge University Press DOI: 10.1017/S1366728904001221

Inflection, thematic roles and abstract Case

JULIA HERSCHENSOHN
Department of Linguistics, University of Washington, Box 354340, Seattle, WA 98195-4340, USA
E-mail: herschen@u.washington.edu

This keynote article proposes a new model of language development based on processing, the sole mechanism of acquisition for the Acquisition by Processing Theory (APT). The language module – adapted from Jackendoff's distinction between integration (building complex structures) and interface (facilitating information transfer at the intersections of the sub-modules) – has phonological, syntactic and conceptual processors that interact with each other, working memory and lexical entries. On the one hand, the application of Jackendoff's model to language learning bridges a gap between syntactic theory and processing. On the other hand, its breadth sometimes brings together strange bedfellows, as, for example, minimalist syntax and APT's 'clear affinities with connectionism' (p. 17). In order to explore this hybrid model within a focused area, I will limit my comments to a discussion of features of I(nflection), Case and thematic (theta) roles, highlighting the question of how much syntactic theory APT needs. I will suggest that the model could profit from a greater exploitation of morphological input and its role in parameter setting.

APT uses a syntactic framework that draws on different theoretical models. Although the authors appeal mainly to Principles and Parameters (P&P), their accounts of verb raising and Case assignment seem based on Chomsky's (1995) Minimalist Program, in which structural Case checking is done in subject (nominative) and object (accusative) agreement nodes through a Spec-Head relationship, and movement is parametrically determined by strength of functional features. For example, verb raising is determined by a strong feature of French I, whereas the weak value of English I entails no verb raising. In contrast, in the P&P model Case is assigned through government (accusative) or Spec-Head agreement (nominative), and verb raising is determined by a positive setting of the Opacity Parameter (Pollock, 1989). In more recent work (Chomsky, 2000, 2001), Case checking may be accomplished through long distance agreement as well as overt movement; feature strength has been abandoned and head movement is disallowed.

Truscott and Sharwood Smith (henceforth T&SS) select pieces from the different models and then combine these ideas with processing, as in this selection concerning the verb raising parameter.

> If it is set at [strong] (the activation level of the latter is higher than that of [weak]), when it becomes part of a representation this setting will have the consequences shown in (2).
>
> (2) *We finished quickly our meal.
>
> Placement of the verb *finished* in the I position, to the left of the adverb *quickly*, results in a sentence that is ungrammatical in English (in contrast to French, for example). Thus, English learners must end up with the [weak] setting, which blocks this structure. (p. 9)

Apparently, hypothesis testing leads learners to attempt raised verbs in English only to discover that the resulting sentences are ungrammatical, a discovery that must be inferred from lack of contrary evidence, since input such as (1) is non-existent. Given that input is the basis for processing, one wonders how the learner could entertain testing a hypothesis that is not provided in the input. Furthermore, evidence from L1A and L2A questions the hypothesis testing idea. L1 learners of English do not raise inflected verbs above negation (Pierce, 1992), while francophone L2 learners of English persist in raising verbs even when provided rich input to the contrary (Trahey and White, 1993).

In contrast to verb raising's feature strength account, T&SS propose that the null subject parameter is determined by a [+pro licensing] feature of I. They are thus inconsistent in the theory they put forth to explain parametric variation, and they do not provide theoretical or empirical justification for this inconsistency. An account that might lend more substantial evidence to their analysis is the notion of morphological richness as a determiner of parameter setting – generally the surface morphology would provide overt interpretable features (helpful as acquisition input) that license the uninterpretable functional features responsible for movement and agreement. In terms of L2A, they could also profitably consider the role of L1 transfer (Schwartz and Sprouse, 1996) and underspecification of features (White, 2003). But finally, it is unclear if they really need the theory to explain the acquisition by processing, since generative models tend to be abstractions that don't relate to real-time processing, while associationist models depend on input alone.

T&SS's fairly abstract account of Case holds that Case is instantiated by a functional category determined by the Case assigning governor (I for nominative, V for accusative). According to APT, language learners use UG – which furnishes a template of functional categories, features and (presumably) principles – to parse incoming data and to extract morphosyntax and phonology. 'In constructing its representations, the syntactic processor always connects each NP to one of these Case items. When no suitable item yet exists (mainly in the early stages of acquisition), it writes $Case_i$ onto an empty node and temporarily connects it to the NP' (p. 10); the new Case is reinforced through frequency, becoming part of the syntactic repertory. The idea that new functional categories can be generated whenever the learner encounters an unknown item is highly unconstrained. If UG already provides the template, then Case nodes already exist and are simply called into play, a scenario more constrained yet problematic in terms of proliferation of functional categories and cross-linguistic variation. Even if Case nodes only relate to structural Case, languages may differ slightly as to the number of distinctive structural Cases in the inventory. It seems that a more morphological account relying on feature specification and agreement would handle both L1 and L2A more easily.

According to T&SS, thematic roles are conceptually linked to Case through probabilistic association during processing. 'When a learner encounters "Pat hit Chris", the syntactic processor is likely to assign nominative Case to the first NP (based on government by I), leading the SS/CS interface processor to activate the CS item coindexed with nominative and connect it to the CS counterpart of "Pat".' (p. 11) This account of Case assignment uses directionality rather than the Case agreement nodes described above to explain the processing of nominative. Moreover, it apparently treats nominative as an interpretable feature, whereas usually Case, as many other functional features, is considered uninterpretable (Chomsky, 1995). It's unclear how Case could have a conceptual counterpart since it is strictly syntactic and lacks conceptual content. The authors suggest that the conceptual counterpart of nominative case is the [agent] theta role derived through many processing events, noting that nominative is associated with agenthood 90% of the time (cf. Figure 4 on p. 10). The authors do acknowledge that the nominative-agenthood contingency is not absolute, as verbs like *receive* and the passive construction with non-agentive subjects show. So they then opt for a probabilistic linking of abstract Case to the most likely theta role, an analysis that might work in terms of processing, but is irreconcilable with the idea of uninterpretable Case. The distribution and nature of theta roles are two phenomena that have been debated for years with a remarkable lack of consensus (cf. Baker, 1988; Grimshaw, 1992; Hale and Keyser, 1993; Herschensohn, 1996). Generative Case theory explicitly rejects a necessary alignment of Case and theta role, particularly in constructions involving non-thematic argument positions (e.g. passive, raising) in which instances of movement and Case are theoretically, not probabilistically, motivated.

The discussion of the Competition Model (Mac Whinney, 2001) following the section on Case and theta roles, gives an illustration of how the APT learning procedure works and sheds further light on Case and theta roles. Speakers weigh cues in order to infer morphosyntax, in English taking word order, but in Italian taking agreement as the best cue. T&SS suggest that 'in English the [left] value has an extremely high level while the [right] value is virtually zero. Thus, I must assign nominative to the NP to its left' (p. 12). Italian shows no directionality, but uses agreement as the cue. Learners use the probabilistic method for determining word order, Case and theta role, linking, for example, nominative case with agenthood. T&SS thus attribute syntactic learning to a stochastic analysis of word order. Once again, they might find a more appropriate account by considering richness of morphology, since the English-Italian difference is due – according to many accounts – to the rich inflection in verbal person-number agreement in Italian as opposed to English. Such an approach also takes inflectional input into consideration as an additional (to word order) clue to the learner. It is difficult to reconcile the probabilistic learning (which could be handled by a connectionist type account) with the abstract system of Case linked to functional categories presented in the earlier section.

Since both L1 and L2 acquisition are considered equivalent by this model, the differential outcomes of the two processes are not explained. T&SS claim that the system of abstract Case is easily established since it is inferred from UG categories, but they 'will not explore the relations between Case items and PS items here, except to note that they can be quite complex and variable, including null PSs, many-to-many mappings, and interactions with other SS items, especially those responsible for gender and number agreement' (p. 10). Surface morphology is not, however, a trivial question. Children learning L1 in the Root Infinitive stage (Wexler, 1994) do not use inflection and Case marking, but they pass this stage to acquire full inflectional ability in a relatively short period. Adults are notorious for lacking inflectional ability in L2, even when they are expert learners (Lardiere, 1998; Prévost and White, 2000; Franceschina, 2001; Herschensohn, 2001); the diagnosis of morphological errors has been a topic of intense discussion in L2 literature.

In conclusion, APT provides a model that forces a consideration of both processing and UG. I submit that the role of morphology is crucial to parametric variation with respect to null subjects, verb raising and word order, and that its integration into the APT model would lead to a better account of how both input and learner production contribute to acquisition through processing.

References

Baker, M. C. (1988). *Incorporation: A theory of grammatical function changing*. Chicago: University of Chicago Press.

Chomsky, N. (1995). *The Minimalist Program*. Cambridge, MA: MIT Press.
Chomsky, N. (2000). Minimalist inquiries: the framework. In R. Martin, D. Michaels & J. Uriagereka (eds.), *Step by step: Essays on minimalist syntax in honor of Howard Lasnik*, pp. 89–155. Cambridge, MA: MIT Press.
Chomsky, N. (2001). Derivation by phase. In M. Kenstowicz (ed.), *Ken Hale: A life in language*, pp. 1–52. Cambridge, MA: MIT Press.
Franceschina, F. (2001). Morphological or syntactic deficits in near-native speakers? An assessment of some current proposals. *Second Language Research*, 17, 213–247.
Grimshaw, J. (1992). *Argument structure*. Cambridge, MA: MIT Press.
Hale, K. & Keyser, S. J. (1993). On argument structure and the lexical expression of syntactic relations. In K. Hale & S. J. Keyser (eds.), *The view from Building 20*, pp. 53–110. Cambridge, MA: MIT Press.
Herschensohn, J. (1996). *Case suspension and binary complement structure in French*. Amsterdam & Philadelphia: John Benjamins.
Herschensohn, J. (2001). Missing inflection in L2 French: Accidental infinitives and other verbal deficits. *Second Language Research*, 17, 273–305.
Lardiere, D. (1998). Case and tense in the fossilized steady state. *Second Language Research*, 14, 1–26.
MacWhinney, B. (2001). The Competition Model: The input, the context, and the brain. In P. Robinson (ed.), *Cognition and second language instruction*, pp. 69–90. Cambridge: Cambridge University Press.
Pierce, A. E. (1992). *Language acquisition and syntactic theory: A comparative analysis of French and English child grammars*. Dordrecht: Kluwer.
Pollock, J.-Y. (1989). Verb movement, universal grammar and the structure of IP. *Linguistic Inquiry*, 20, 365–424.
Prévost, P. & White, L. (2000). Truncation and missing inflection in second language acquisition. In M. A. Friedmann & L. Rizzi (eds.), *The acquisition of syntax*, pp. 202–235. London: Longman.
Schwartz, B. D. & Sprouse, R. (1996). L2 cognitive states and the Full Transfer/Full Access model. *Second Language Research*, 12, 40–72.
Trahey, M. & White, L. (1993). Positive evidence and preemption in the second language classroom. *Studies in Second Language Acquisition*, 15, 181–139.
Wexler, K. (1994). Optional infinitives, head movement and the economy of derivations. In D. Lightfoot & N. Hornstein (eds.), *Verb movement*, pp. 305–350. Cambridge: Cambridge University Press.
White, L. (2003). *Second language acquisition and Universal Grammar*. Cambridge: Cambridge University Press.

Parameters or cues?

BRIAN MACWHINNEY
Department of Psychology, Carnegie Mellon University, Pittsburgh, PA 15213, USA
E-mail: macw@cmu.edu

Truscott and Sharwood Smith (henceforth T&SS) attempt to show how second language acquisition can occur without any learning. In their APT model, change depends only on the tuning of innate principles through the normal course of processing of L2. There are some features of their model that I find attractive. Specifically, their acceptance of the concepts of competition and activation strength brings them in line with standard processing accounts like the Competition Model (Bates and MacWhinney, 1982; MacWhinney, 1987, in press). At the same time, their reliance on parameters as the core constructs guiding learning leaves this model squarely within the framework of Chomsky's theory of Principles and Parameters (P&P). As such, it stipulates that the specific functional categories of Universal Grammar serve as the fundamental guide to both first and second language acquisition. Like other accounts in the P&P framework, this model attempts to view second language acquisition as involving no real learning beyond the deductive process of parameter-setting based on the detection of certain triggers. The specific innovation of the APT model is that changes in activation strength during processing function as the trigger to the setting of parameters. Unlike other P&P models, APT does not set parameters in an absolute fashion, allowing their activation weight to change by the processing of new input over time. The use of the concept of activation in APT is far more restricted than its use in connectionist models that allow for Hebbian learning, self-organizing features maps, or back-propagation.

Like other minimalist models, APT model assigns a central role to a universal processor and attributes all language variation to syntactic frames stored inside the lexicon. Chomsky's shift to minimalism brings his theory in line with two consistent emphases in the Competition Model – the centrality of the processor in learning and the encoding of structure on lexical items. The idea that the processor itself is universal was central to the first version of the Competition Model (Bates and MacWhinney, 1982). Since 1978, I have also argued that learning of L1 syntactic classes begins on the level of the individual item and that higher-level syntactic patterns emerge from the gradual formation of lexical groups. Although transfer effects complicate matters in the case of L2 acquisition, the Competition Model continues to develop based on the principles of lexicalist, bottom-up learning organized about a universal processor.

T&SS seem to be picking up on some of these earlier Competition Model accounts in their brief descriptions of parameter-setting within the processor. For example, they talk about identifying the word 'horse' inside the utterance 'The horse is beautiful' (p. 7). As I have argued in perhaps a dozen articles, this type of learning occurs under the pressure of the processing of item-based constructions that open up positional expectations for particular meaningful elements. In this particular example, the relevant frame is the item-based construction based on the word *the* that opens up a following slot for a common noun. The definite article does not have a clear expectation for a count noun as opposed to a mass noun, but it will not tolerate a proper noun, adverb, or verb. Thus, it provides a highly valid cue to the extraction of the meaning of the following word.

Given this apparent agreement on the facts of learning, it was with some surprise that I read the claim from T&SS that 'the Competition Model posits two distinct types of entities, formal characteristics and processing strategies that use them; each must be acquired' (p. 16). This interpretation seems to reflect a misunderstanding of the Competition Model. In our model, all learning is situated within the lexical frames of individual items. Although Competition Model experiments lump together groups of verbs for the purposes of experimental control and generalizability, the theory has always assumed, and specific studies have consistently demonstrated, that the locus of learning is in the valency frames of individual predicates. The Competition Model assumes that the processor is universal and unlearned. This idea is taken to be a new concept in minimalism, but it has been fundamental to the Competition Model since 1982.

Despite our agreement on the central role of a universal processor, the ATP model differs radically from the Competition Model in its view of what is learned. For ATP, there is a small set of universal functional categories that are innately available to the processor. During processing, those features that are used by a language are activated to a certain level and that is all that is necessary to account for language acquisition. The Competition Model also assumes that acquisition occurs during processing. However, unlike ATP and P&P, it assumes that each predicate has its own unique semantic configuration of roles. For example, the verb *give* has three arguments that include a giver, a recipient, and an object transferred. The verb *chop* has two arguments that include a chopper and an object chopped. We insist on this low-level representation of syntactic arguments because of evidence regarding the item-based nature of early predicate learning, particularly verb learning, as well as the impossibility of establishing a universal typology of semantically grounded cases. However, we also recognize that young children and beginning second language learners are soon able to generalize across verb frames to extract language-specific higher-order syntactic categories such as subject or instrument. When learners come to form these higher-order categories, they rely on the concepts of perspective and embodiment. The categories organized by perspective emerge not from an autonomous linguistic module, but from general properties of human cognition.

Two decades of work in the P&P framework have failed to yield a uniform universal account of first language acquisition based on parameter-setting. On the other hand, bottom-up, statistical learning accounts, such as the Competition Model, have made continual progress in extending our understanding of the details of L1 learning. However, it might be the case that ATP would still be of some value as an account of second language acquisition. In this area, T&SS rely primarily on data from studies of adverb placement by French learners of English as a second language. The evidence here is that even advanced learners have problems learning to place the adverb before the verb. But, if learners are hearing examples of English constructions with adverbs before the verb, should not this input lead to a retuning of the weight of the feature that determines preverbal placement? Apparently, T&SS hope to solve this problem by suggesting that the errors are due not to faulty L2 learning, but to code-switching from L1. Presumably, they would argue that each French adverb and verb has a frame that was set to an activation level that placed it in preverbal position. When the French speaker comes to producing English, these French lexical frames become activated, despite the fact that the actual words being produced are in English. If ATP is making this claim, then I would say that what T&SS call code-switching is nothing more than a terminological equivalent to what the Competition Model calls lexically-based transfer. But if parameters are in turn lexically based, don't we have a fundamental logical problem? If minimalism links functional categories to specific lexical items, it vitiates its own claim for their innateness and universality.

It should be possible to examine the exact nature of lexically-based transfer empirically. To begin, we should note that English itself seems to belie the descriptive utility of the single parameter account when it permits sentences such as 'I like to eat sometimes Indian and sometimes Chinese', 'Jim likes always something different in his cereal' or 'You go how often to the store?' If English (or any other language for that matter) were strictly governed by parameters, we would never see the complexity and variation that we really have in syntax. In order to produce the complete set of predictions for lexical transfer effects in adverb placement, one would also have to examine possible cracks in the French system to see where they match up or fail to match up with English. Constructing such a complete contrastive analysis is beyond my ability in French and far beyond the scope of this commentary, but it would be the best way to explore the ATP model empirically. My guess is that, once the contrastive analysis is produced and the relevant studies are run, the results will demonstrate that learners are quick to pick up consistent L2 cues, and that it is inconsistency in L2 that opens up the door for transfer from L1. I hope that T&SS continue to develop their ideas in ways that would eventually lead to such empirical tests.

References

Bates, E. & MacWhinney, B. (1982). Functionalist approaches to grammar. In E. Wanner & L. Gleitman (eds.), *Language acquisition: The state of the art*, pp. 173–218. New York: Cambridge University Press.

MacWhinney, B. (1978). The acquisition of morphophonology. *Monographs of the Society for Research in Child Development*, 43 (1), 1–123.

MacWhinney, B. (1987). The Competition Model. In B. MacWhinney (ed.), *Mechanisms of language acquisition*, pp. 249–308. Hillsdale, NJ: Lawrence Erlbaum.

MacWhinney, B. (in press). Rethinking the logical problem of language acquisition. *Journal of Child Language*, 30.

Processing perspectives in SLA research and their compatibility

MANFRED PIENEMANN
University of Paderborn, Germany
E-mail: Pienemann@HRZ.upb.de

Truscott and Sharwood-Smith's (henceforth T&SS's) paper offers an interesting set of hypotheses about one possible processing perspective in research on language acquisition. What is striking about this exposition of their model is that it ignores almost entirely the context of previous research on this issue. Embedding their exposition in its historical context would have in no way diminished the innovation entailed in their proposal. Instead it could have served as an instrument for ensuring conceptual parsimony, which in my view is flawed in this exposition.

Language processing has been a key aspect of SLA (and L1) research right from the beginning. For many of us this is marked by Corder's (1967) paper, which included, among other things, a discussion of input processing that was taken up by other scholars. Other lines of research included work on procedural skills (e.g. Levelt, 1978; McLaughlin, Rossman and McLeod, 1983; McLaughlin, 1987), work on operating principles (Andersen, 1984; Slobin, 1973), work on processing strategies (e.g. Clahsen, 1984) and my own work on processability theory (Pienemann, 1998a, b, 2003). A brief overview is available in Pienemann (2003). Further relevant approaches are described in the 2002 Special Issue of *Second Language Research*.

In all of these lines of research, language processing served as an explanatory construct for a number of issues in second (and partly in first) language acquisition. These included the origin of linguistic knowledge (e.g. Bates and MacWhinney, 1981; Andersen, 1984), L2 developmental trajectories (e.g. Pienemann, 1998a, b) and inferential mechanisms (e.g. Carroll, 1999, 2001). Depending on their orientation in the philosophy of science, these authors developed their approaches either as modular or as general-cognitive explanatory constructs. A great deal of the debate surrounding these approaches has been directed at two key issues: (1) procedural explicitness and (2) representation of linguistic knowledge. It may be useful to sketch out some of this discussion in order to demonstrate its repercussions for the approach presented in T&SS's paper.

One line of research that offers a processing perspective on L2 development focuses on L2 procedural skills. McLaughlin assumes that '[t]o learn a second language is to learn a skill' (McLaughlin, 1987, p. 133) and that L2 learning 'requires the automatisation of component sub-skills' (ibid.). Similarly, other authors have also expressed the view that language acquisition entails the acquisition of procedural skills (e.g. Levelt, 1978; Hulstijn, 1990; Schmidt, 1992).

One early approach to SLA that incorporates a processing perspective is Krashen's (1985) 'monitor model', which received ample attention in the 1980s and has been subjected to extensive critiques (e.g. McLaughlin, 1978; Gregg, 1984; Long, 1985). In his 'input hypothesis' Krashen claims that language is acquired by learners receiving 'comprehensible input'. The 'input hypothesis' is aimed at explaining two things, namely (1) the inferential mechanisms that drive the acquisition process and (2) the assumed universal order of acquisition. Critics point out that the 'input hypothesis' cannot be operationalised for any of its components. As a result, it cannot be tested empirically. In other words, Krashen's model evades the issue of specifying the architecture of the L2 processor and the inferential mechanisms involved.

Van Patten's (1996) work is an example of later mainstream research on input processing. It follows the main idea of the 'input hypothesis' and stipulates two sets of input processing strategies in an attempt to spell out aspects of the architecture of the L2 processor that regulate attention and the assignment of semantic roles. Van Patten follows Corder (1967) in distinguishing between 'input' and 'intake' and stipulates attention as the necessary condition for input to be transformed into intake. This approach is limited to one domain of language processing and is unable to generate the data it is designed to explain.

Carroll (1999, 2001) reviews the literature on L2 input processing and concludes that the standard assumption, based on Corder's (1967) input–intake distinction, according to which 'perception is regulated only by attention, which in turn is regulated by intention' (Carroll, 1999, p. 343), is not supported by any explicit theory of attention. Carroll proposes the Autonomous Induction Theory, which is an explicit theoretical framework for the induction of linguistic representation from linguistic input. Her position is compatible with a modular view of processing and a UG-position on cognition, and is thus juxtaposed to the functionalist orientation of the standard view on the attention filter in processing. This work highlights the enormity of the task of specifying the inferential mechanisms that explain how input becomes intake.

Processability Theory (PT; Pienemann, 1998a, b) is a modular processing approach that is based on Levelt's (1989) approach to language production. Within this framework the architecture of the language processor is modelled as a hierarchy of processability that is applied to developing linguistic systems. The logic underlying PT is the following: at any stage of development, the learner can produce only those L2 linguistic forms which the current state of the language processor can handle. The processability hierarchy is operationalised through an implementation in Lexical-Functional Grammar. This implementation of the processability hierarchy into an LFG-based description of a given language affords us a formal and testable prediction of developmental trajectories in L1 and L2 learners. In this way PT achieves procedural explicitness, and it adds a processing dimension to learnability, which is traditionally defined as a purely logical-mathematical problem (e.g. Berwick and Weinberg, 1984). In sum, PT is procedurally explicit, cross-linguistically valid, supported by extensive empirical studies, and it is compatible with an LFG account of linguistic knowledge.

The 'competition model' (Bates and MacWhinney, 1981, 1982) is a further approach to language acquisition that offers a processing perspective. It is a functionalist approach that is based on the assumption that linguistic behaviour is constrained, among other things, by general cognition (and not by a language-specific cognitive module) and communicative needs. Following the functionalist tradition, Bates and MacWhinney assume that 'the surface conventions of natural languages are created, governed, constrained, acquired, and used in the service of communicative functions' (Bates and MacWhinney, 1981, p. 192). According to this model, it is the task of the language learner to discover the specific relationship between linguistic forms of a given language and their communicative functions. In the competition model, the process of learning linguistic forms is driven by the frequency and complexity of form-function relationships in the input. In this context, the majority of L2 learning problems is modelled in connectionist terms.

The connectionist framework utilised in the competition model has been subjected to severe criticism, as evident in Pinker and Prince (1987), who analysed the developmental assumptions of connectionist models and falsified them manifold in empirical studies. Whereas connectionists assume that linguistic rules are 'merely convenient approximate fictions' (Pinker and Prince, 1987, p. 1), these authors conclude that 'connectionists' claims about the dispensability of rules in explanations of the psychology of language must be rejected, and that, on the contrary, the linguistic and developmental facts provide good evidence for such rules' (ibid.).

In this context the reader is surprised to find that in T&SS's paper key features of the connectionist architecture are grafted onto an account of linguistic knowledge that is based on Principles and Parameters and minimalism. In the view of Pinker and Prince, symbolist and connectionist accounts of language acquisition are in principle irreconcilable. If one does not share this view (cf. Hulstijn, 2002, for a unified position) one would expect a detailed justification of the grafting proposal at an epistemological and a technical-descriptive level. At this stage it is difficult to see the exact detail of such a justification, because the nature of the graft remains unclear. This is mainly due to the fact that T&SS do not specify the assumed parsing algorithm in an operationalisable manner (apart from providing illustrative examples). This makes it difficult to evaluate how the output of the processor can be translated into the categories found in the assumed system of linguistic representation.

This lack of procedural explicitness also makes it difficult to test the proposal beyond the examples given or to apply it cross-linguistically. What is described of the architecture of the grammatical encoder is mainly based on assumptions and no empirical data are offered in support of the proposed model of language processing. These last two paragraphs may at first glance be asking a lot of a new theory of language acquisition. However, we cannot compromise on the core requirements of any such theory:

- procedural explicitness,
- theoretical parsimony,
- cross-linguistic validity and
- empirical soundness.

Certainly, many approaches to language acquisition do not satisfy all of the above requirements, and there are some recent ones that do explicitly attempt to overcome this shortcoming (including Carroll, 2001). In my view, this shows that constructing a theory of (first and second) language acquisition (and processing) is a tall order indeed. T&SS's paper certainly provides an interesting point of departure for such an undertaking.

References

Andersen, R. (1984). The one-to-one principle of interlanguage construction. *Language Learning*, 34, 77–95.

Bates, E. & MacWhinney, B. (1981). Second-language acquisition from a functionalist perspective: Pragmatic, semantic, and perceptual strategies. In H. Winity. (ed.), *Native language and foreign language acquisition*, pp. 190–214. New York: Annals of the New York Academy of Sciences 379.

Bates, E. & MacWhinney, B. (1982). Functionalist approaches to grammar. In E. Wanner & L. R. Gleitman (eds.), *Language acquisition: The state of the art*, pp. 173–218. Cambridge: Cambridge University Press.

Berwick, R. & Weinberg, A. (1984). *The grammatical basis of linguistic performance: Language use and language acquisition*. Cambridge, MA: MIT Press.

Carroll, S. E. (1999). Putting 'input' in its proper place. *Second Language Research*, 15(4), 337–388.

Carroll, S. E. (2001). *Input and evidence: The raw material of second language acquisition*. Amsterdam: John Benjamins.

Clahsen, H. (1984). The acquisition of German word order: A test case for cognitive approaches to second language acquisition. In R. Andersen (ed.), *Second languages*, pp. 219–242. Rowley, MA.

Corder, S. P. (1967). The significance of learners' errors. *International Review of Applied Linguistics*, 5, 161–170.

Gregg, K. (1984). Krashen's Monitor and Occam's Razor. *Applied Linguistics*, 5, 79–100.

Hulstijn, J. (1990). A comparison between information-processing and the analysis/control approaches to language learning. *Applied Linguistics*, 11, 30–45.

Hulstijn, J. (2002). Towards a unified account of the representation, processing and acquisition of second language knowledge. *Second Language Research*, 18 (3), 193–223.

Krashen, S. (1985). *The Input Hypothesis: Issues and implications*. London: Longman.

Levelt, W. J. M. (1978). Skill theory and language teaching. *Studies in Second Language Acquisition*, 1, 53–70.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Long, M. H. (1985). A role for instruction in second language acquisition: Task-based language training. In K. Hyltenstam & M. Pienemann (eds.), *Modelling and assessing second language acquisition*, pp. 77–100. Clevedon: Multilingual Matters.

McLaughlin, B. (1978). The Monitor model: Some methodological considerations. *Language Learning*, 28, 309–32.

McLaughlin, B. (1987). *Theories of second language learning*. London: Edward Arnold.

McLaughlin, B., Rossman, T. & McLeod, B. (1983). Second language learning: An information-processing perspective. *Language Learning*, 3, 135–157.

Pienemann, M. (1998a). *Language processing and second language development: Processability Theory*. Amsterdam: John Benjamins.

Pienemann, M. (1998b). Developmental dynamics in L1 and L2 acquisition: Processability Theory and generative entrenchment. *Bilingualism: Language and Cognition*, 1(1), 1–20.

Pienemann, M. (2003). Language processing capacity. In C. Doughty & M. H. Long (eds.), *Handbook of second language acquisition*, pp. 679–713. Blackwell: Oxford.

Pinker, S. & Prince, A. (1987). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition (Occasional Paper, 33). Centre for Cognitive Science, MIT.

Schmidt, R. (1992). Psychological mechanisms underlying second language fluency. *Studies in Second Language Acquisition*, 1, 357–385.

Slobin, D. I. (1973). Cognitive prerequisites for the development of grammar. In C. A. Ferguson & D. I. Slobin (eds.), *Studies of child language development*, pp. 175–208. New York: Holt Rinehart and Winston.

Van Patten, B. (1996). *Input processing and grammar instruction in second language*. Norwood, NJ: Ablex.

Bilingualism: Language and Cognition 7 (1), 2004, 40–41 © 2004 Cambridge University Press DOI: 10.1017/S1366728904001257

Can the APT be a theory of Multiple Grammars and a minimal theory of parsing?

TOM ROEPER
Department of Linguistics, University of Massachusetts, Amherst, MA 01003, USA
E-mail: roeper@linguist.umass.edu

This essay by Truscott and Sharwood-Smith is a valiant attempt with a laudable goal. It seeks to show how different perspectives and disciplines can capture what is happening in acquisition and notably in L2 acquisition. Nonetheless, I think that the results are much closer to an elaborated grammatical theory than an elaborated processing theory (thus it seems less important to me than to the authors to choose one label over the other). Their essential, sensible idea is that where different grammars provide independent analyses of constructions, both are computed (even where the lexical items belong to only one grammar). Therefore conflict is experienced which produces computational demands. However this is pure grammar – just a more sophisticated situation where, as I like to argue, the L2 person has Multiple Grammars (Roeper, 1999), creating more computational work as well. In this respect, the situation is no different from L1 acquisition, where the course of acquisition involves generation and maintenance of multiple grammars, some of which are shed and some retained in the Final State.

The first half of the essay demonstrates enormous erudition and efforts to synthesize, but noticeably does not result in many specific analyses of sentences. It remains at the abstract level. This is a natural prolegomena, but not a deductive system, which is perhaps too much to ask. Nevertheless, it does not seem that we see the promised interaction at work beyond what can be captured by the following simpler system.

Let us suppose that the parsing system outside of grammar consists only of Merge. One element must merge with another, either new or already analyzed. Two questions arise: which dominates which, and what level does Merge engage if one object is complex? These are reminiscent of how MINIMAL ATTACHMENT and LATE CLOSURE decide where to attach (Frazier, 1978). Now we can (possibly) exercise a default assumption: merge high or merge low without engaging language-specific grammatical assumptions. However, as soon as we begin to decide the dominance question, we enter grammatical territory: what categories dominate the mergeable element? The set of available categories will be sufficiently detailed that they vary from grammar to grammar. Modals function like main verbs in German and can be recursive. One must know from a grammar when a modal arises AND know if the grammar gives them a feature that allows recursion. Therefore the grammar is now steering a very elementary parser. And it is important to note that the properties of the parser constrain the set of UG grammars, so one might want to say that the parser also steers the grammar. An example may help.

Suppose one is acquiring a second grammar (English), but the first is assumed to be in the background (German). Then we will assume that both grammars provide an analysis wherever possible. If one heard a sentence like (1),

(1) We should can must and then see if we were able to sell it.

the English speaker will be able to recognize, perhaps slowly, that in a certain context it makes sense: Hollywood has a hard time creating old musty houses. If we could can must and sell it, we would have a big business. In German, one can have several modals, including *can must* (*er kann singen müssen*), and therefore the German speaker would have the extra task of coping with the notion that there are two modals (which does occur in English dialects *I might could do that*). We therefore predict that disambiguation for the German speaker would be significantly harder. There is nothing more involved than saying that the speaker entertains two grammars, maintains those analyses until they are resolved and that such efforts require computational space.

Such effects are sufficiently strong and overt that they are conscious. Presumably much the same happens unconsciously and the theory of Carroll (2001) and Truscott and Sharwood-Smith is aimed in that direction and has inspired these ruminations.

Consider the following sentence in Spanish:

(2) Every student thinks he knows
cada uno de los estudiantes piensa que __ sabe
how he thinks he can best tie
como __ cree que __ puede atar
his own shoes.
sus zapatos mejor

We know that the empty categories in Spanish alternate with filled pronouns. The empty categories are preferentially bound to *every student*, while the full ones are preferentially linked to independent reference. The English speaker comes with a grammar where the full pronouns have both readings, and no empty subjects exist. Therefore, for the version with full NPs, the English speaker has his priorities the opposite of the Spanish speaker when a quantifier is present. This should produce a slight effect.

Much more complicated situations arise. One can allow external reference for the empty NP in Spanish as well, and it may continue down a sequence of empty phrases, such as the sentence we just gave. However, it should be very difficult for the English speaker to allow this reading, since it would be naturally captured in English by full pronouns no matter what.

In Spanish the empty pronouns cannot switch back and forth between distributive and referential. That is, one cannot say:

(3) Here is the teacher. Every student thinks __ will expect that does __ homework.

where the first gap is 'the teacher' and the second is bound to 'the student'. In English, however, it is possible to switch, but in a limited way:

(4) Every student thinks he will assign so much homework that he cannot do it.

This can have a reading where the first *he* refers to the teacher and the second *he* refers back to the student again. Now here is where, if both grammars are engaged for a Spanish speaker, we would predict quite a bit of interference over the long-distance of this sentence. That is, it should be harder for the Spanish speaker to have shifting full pronouns in English, because the Spanish speaker tries to analyze it as if it were non-shifting empty pronouns (thanks to Luis Alonso-Ovalle for judgments and advice).

These examples support the spirit of the Truscott-and-Sharwood Smith enterprise but the method seems a bit different. The notion that access to Multiple Grammars is needed immediately suggests that new processing demands should will arise, which should lead directly to predictions for numerous challenging examples like those we have just discussed. Truscott and Sharwood Smith's work serves as an inspiration to study such L2 effects carefully.

One final question: does the syntax/semantics interface show the same interactions? We could show, of course, that lexical ambiguity will cause interference. Can we also show that more sophisticated interface effects are present (like grammatically varied representations of aspect)? It seems, perhaps too pessimistically, that when the larger realm of human cognition is engaged – where much of our thinking is rather slow (humans ruminate on things for weeks as well as milliseconds) – that we do not yet know how to isolate what functions in the rapid mental computations around language. In other words, basic insights into restricting the interface are still missing. Nevertheless, all attempts – like those of Truscott and Sharwood Smith's – to achieve a synthesis across domains help us to enliven the questions.

References

Carroll, S. E. (2001). *Input and evidence: The raw material of second language acquisition*. Amsterdam: John Benjamins.

Frazier, L. (1978). *On comprehending sentences: Syntactic parsing strategies*. Ph.D. dissertation, University of Connecticut.

Roeper, T. (1999). Universal bilingualism. *Bilingualism: Language and Cognition*, 2 (3), 169–186.

AUTHORS' RESPONSE

How APT is your theory: Present status and future prospects

JOHN TRUSCOTT
Department of Foreign Languages and Literature, National Tsing Hua University, Hsinchu, Taiwan
E-mail: truscott@mx.nthu.edu.tw
MIKE SHARWOOD SMITH
Heriot-Watt University, Edinburgh

To begin with, we thank the commentators for their remarks, which both intrigued and challenged us in various ways. Time and space limits prevent us from doing justice to all the points they raised, so we will focus on those we consider most significant and are best prepared to deal with. One of these points, of course, is the overall character and status of our proposal. Pienemann terms Acquisition by Processing (APT) 'an interesting point of departure' (p. 38) for the development of a theory of acquisition and processing. Harrington and Carroll say in effect that it can't be done; others seem more open to the possibility. In any case, we quickly accept Roeper's suggestion that what we have presented is more a 'prolegomena' (p. 21) to a theory than a full-fledged theory in itself.

We think that misunderstandings have arisen about our putative 'mix and match' or 'hybrid' approach. The point is that APT is, to some extent, applicable to different approaches. More specifically, we found that the best way of introducing APT was by taking approaches that would be familiar to readers and then showing what possible consequences an APT approach would have for them. For example, our discussion about how two theoretical linguistic frameworks, Principles and Parameters and Minimalism, might incorporate APT is not the same as patching together a hybrid theory of everything (pace Carroll) or, indeed, creating a mainly grammatical theory (pace Roeper). We return to these points below.

In what follows, we look first at the commentators' discussion of APT as an account of language development, its core role. We then turn to its other aspects – syntactic theory and processing theory – and then consider compatibility issues in more depth. We conclude with a consideration of future directions for research, based on commentators' suggestions.

APT and language development

Our proposal is, above all else, an account of language development, so the natural starting point is provided by the many comments on this topic. Because the removal of LAD is at the heart of APT, we begin with this issue before turning to some more specific points discussed in the commentaries.

Ironically, our framework began as an attempt to formalize a theory of LAD, but its monitoring and corrective role simply dissipated as we developed our proposals further: The application of Occam's razor more or less proposed itself. The mechanisms needed to explain the nature and use of language appeared to be adequate to explain its development as well. Hence the concept of LAD survives only as an abstraction from UG-based processing operations.

Bickerton is puzzled by our rejection of LAD, because he considers it simply UG seen from another angle. But 'LAD' carries theoretical baggage that is not inherent in the idea of UG and which, we have argued, has no place in it. This baggage consists of a set of explicit learning mechanisms and principles, existing specifically for the purpose of learning. Our claim that nothing of this sort exists is one which the typical UG theorist would not consider obvious and uncontroversial, to say the least. We obviously share Bickerton's view that 'Processing with the aid of UG seems the only way to go' (p. 21). But this is by no means a commonsensical statement expressing ordinary views of UG and LAD. If the quoted statement does express his position then, frankly, he should welcome our proposal as a long-overdue effort to straighten out pervasive misconceptions.

Carroll rejects our claim that APT is a parsimonious approach, but her argument largely misses the point. An argument against our position would seek to show that we have a level of complexity substantially beyond that which is needed to explain bilingual knowledge and use of language, including the processing done by learners, which inevitably involves very substantial gaps in the knowledge bases. But Carroll does not even try to do this. As a result, her argument is sketchy at best and irrelevant at worst.

MacWhinney suggests that we misunderstand learning in the Competition Model. This may well be true. Our discussion was based on several standard descriptions of the CM (Bates and MacWhinney, 1982, 1987, 1989; MacWhinney, 2001). In these sources we do not find the lexical learning that he now describes ('all learning is situated within the lexical frames of individual items'

(p. 35)). Nor do we find the 'universal and unlearned' processor (p. 35). What we do find is descriptions of how English speakers, for example, assign agency in a sentence, without reference to the lexical content of the sentence; how this process differs across languages; and how children learning different languages do or do not develop certain strategies for assigning agency. Maybe these are actually high-level ways of referring to purely lexical learning and the workings of an invariant processor, but this is not stated, and the translation is far from obvious.

Our confusion is made worse by apparently conflicting statements in MacWhinney's commentary. After stating that 'all learning is situated within the lexical frames of individual items', he then says that learners 'generalize across verb frames to extract language-specific higher-order syntactic categories' (p. 36). Why is this not a contradiction? And what other generalization processes does the model assume beyond the learning that occurs within lexical frames?

Part of the issue is how English speakers assign agency in a case like 'The lion glurps the tiger'. The typical speaker will have no doubt that the agent role should be assigned to 'the lion', despite the fact that 'glurp' has no lexical frame. If learners generalize from individual verb frames to higher-order categories, do they also generalize from individual verb frames to a principle that preverbal NPs are very likely to be agents (establish a cue strength for word order that is independent of any particular lexical frame)? The sources cited above certainly give the impression that abstract principles of this sort exist and are responsible for assignment of agency. If this is the case, then the CM does hypothesize an additional type of entity, which must also be acquired.

We recognize, of course, that we do not have in-depth knowledge of the Competition Model and therefore may have misunderstood things. Whether or not this is the case, we hope that further discussion will clarify the point.

Turning to some more specific issues regarding language development in APT, we find several cases of misunderstandings in the commentaries. Carroll's comments on our discussion of child–adult differences in acquisition suggest an extremely impoverished view of our proposals. We discussed input competition, differences in working memory, and output competition as possible explanations for these differences, and noted that the answer could well lie in a combination of these and other factors. Carroll's rebuttal is that differences in working memory are not an adequate explanation.

Her other comments on acquisition tend to be sketchy, and we often find them hard to follow. At one point, she seems to be saying that, according to APT, learners can only combine two items in production if they have encountered that particular combination in comprehension. This would be an absurd position and we do not know why she thinks we are obliged to adopt it. No less confusing is her claim that (because we reject LAD, apparently) we cannot possibly deal with differences in the way early and late bilinguals use gender marking in comprehension (p. 23; Guillelmon and Grosjean, 2001).

Dijkstra and Haverkort say that language development in APT includes the UG-guided 'formulation of rules' in response to repeated input (p. 27).We see this as a description at a very high level of abstraction, which we prefer to avoid in this context. What is actually happening, we suggest, is that items are being written in the lexical stores, based on the application of UG constraints to input, and are then undergoing gradual changes in their resting activation levels through further processing activity. Throughout, their use is also constrained by UG. When -*ed* is first processed as a past tense item, for example, it is written in the stores as such. Because it is written as a tense form, UG constraints force it to appear in the structural position T (Tense) when it enters an SS representation. Its appearance in this position then ensures that it will attach to verbs. Its subsequent use with verbs in production will be contingent on its activation level, both absolutely and relative to any other potential occupants of the position. Through repeated use, the level will rise and so the affix will routinely appear in the appropriate position. One could see this as the development of a rule, that -*ed* attaches to verbs to express past tense, but we consider this way of looking at the process more misleading than helpful.

The situation is similar for Herschensohn's discussion of 'hypothesis testing' in APT. We have no such concept. When the syntactic processor constructs a representation during comprehension, it inserts I. If it cannot construct a legitimate representation for the current input using the currently dominant feature value, it seeks to raise the other value enough for it to temporarily take over. Every time this alternative value appears in the ultimate representation, its resting activation level will be slightly raised. The result is that one value will ultimately attain so high a level that it can no longer be reversed. Throughout, whichever value has so far proven most useful, most often, controls production. Calling this process hypothesis testing is, again, more misleading than helpful. Regarding Herschensohn's point that indirect negative evidence is needed, we already pointed out possible positive evidence (for the processors) and raised the possibility that the English setting is the default, obviating the need for evidence.

Finally, Dijkstra and Haverkort suggest that language development in APT strongly resembles Dynamic Systems Theory (e.g. Edelman, 1992). We have not had time to seriously consider the issue, but the point could be quite important.[1]

APT and syntactic theory

In discussing the role of syntactic theory, we should first emphasize once again that our proposal at this point is best seen as a prolegomena rather than a full-fledged theory. The particular linguistic analyses we presented were thus intended as illustrations of the logic and as demonstrations that an approach of this sort is feasible. Some of the commentators pointed out inconsistencies in our selection of syntactic approaches, which we acknowledge. This does not imply the advocacy of a theoretically hybrid model. The principles of APT can be applied to a (not unlimited) variety of linguistic models, but the full working out of such applications and the choice of a most promising one we leave for future research.

So we agree with some of the commentators that in a fully developed APT model the linguistic assumptions need to be more explicit, with a more clearly coherent view of syntax. But we take issue with several related points that were made. Most disturbingly, there is serious confusion about the nature of our approach.

Carroll believes that we are proposing a Minimalist theory, with all its devices and all its constraints. A similar misunderstanding is found (inconsistently) in MacWhinney's comments, though it does not play so large a role there. We had hoped we had made it clear that this was definitely not what we were doing. We have adopted the Minimalist idea that variation is restricted to the lexicon, but our use of the approach does not go far beyond that and we do make extensive use of Principles and Parameters. This seemed a better strategy than adopting what might have been a sensible option, namely Construction Grammar, this being preferred by Jackendoff (2002), since it would then be more difficult to explain the principles of APT in terms familiar from mainline acquisition literature. Jackendoff himself adopts a similar open-ended presentation strategy in explaining his model in Jackendoff (1997).

In MacWhinney's discussion of English and French word order he says we presumably hold the position that frames of individual verbs and adverbs are responsible for the ordering of adverbs relative to verbs and for the problems French speakers have in getting the English order right. From this premise he infers a logical problem, that this dependence on lexical factors ('lexical' in the traditional sense) makes parameters meaningless (p. 000). We were quite explicit about this in our discussion of I: we follow familiar linguistic accounts in attributing the phenomena to a feature of I, not to individual verbs or adverbs. This feature, with its [strong] and [weak] values, is a parameter.

APT and processing

Our proposal is also intended to be an account of processing. This aspect also drew some interesting comments.

Bickerton's main theme is that we, along with many others, failed to recognize the centrality of syntax. In particular, he wants to minimize the role of the conceptual processor. But his arguments assume that a great deal of the burden of processing is carried by conceptual factors, including context, decisions about whether an analysis is 'gibberish', likely referents of an expression, and apparently the meanings of words. He also seems to assume that they have nothing to do with the conceptual processor, arguing that because these factors plus the syntactic processor can carry the load of processing, the conceptual processor has little to do. So his argument is apparently that if one assumes that much of the content we attribute to the conceptual processor is actually located somewhere else in the processing system (where?[2]), then the narrowly defined conceptual processor that results has only a limited role. This may well be true, but we do not see any interesting implications. We also note that when he discusses syntactic factors involved in the rejection of unacceptable analyses, he is elaborating on our account, not challenging it. We said both syntactic and conceptual factors are involved.

Some of Harrington's comments on syntactic priming research are helpful. In particular, he raises serious questions about whether the effects really have any implications for APT (and for processing theory in general), based on their sometimes surprising duration. We find his interpretation of Bock and Loebell's (1990) work less helpful. In particular, his statement that it showed priming effects from a rejected analysis is doubtful (p. 30). The authors' own conclusion was quite different: when subjects' processing of a locative sentence increased their likelihood of subsequently producing passives, this effect resulted directly from structural similarity between the locative and the passive, not from the characteristics of any rejected analysis. And it is doubtful that a passive analysis was ever produced for the locative sentences used in the study, as those sentences all involved progressive verb forms (e.g. 'The 747 was landing by the control tower'), in which an incremental processor could rule out a passive reading before the *by*-phrase was reached. The test case we suggested was designed to ensure that the verb

[1] An anonymous reviewer also raised the point, though too late in the process for us to accommodate it.

[2] The discussion is especially confusing because Bickerton has prior context, a phone call for example, influencing the comprehension of a sentence 'long before recourse to any conceptual processor' (p. 21). In a Jackendovian model like ours, this would mean the knowledge that a phone call was just made is either phonological or syntactic.

form was consistent with a passive structure. Finally, the experimental design does not appear to rule out alternative explanations for the effects, such as priming of *by* itself. The weakness of the effect might fit such an explanation.

We are not sure what Harrington has in mind when he discusses descriptive and explanatory uses of ‘activation’(p. 30). It looks as if he is saying that the notion can only have explanatory value if we use it as a literal claim about brain function (we do not). Such a view would be difficult to understand.

Carroll’s assertion that we cannot possibly deal with cross-linguistic variation in relative clause attachment preferences (e.g. Cuetos and Mitchell, 1988) seems to be based, again, on an impoverished view of APT. Her argument is that the variation is not lexical (i.e., it is based on general syntactic structure) and therefore cannot be handled by APT. On the traditional notion of ‘lexical’ her premise is probably correct. But, our notion of ‘lexical’, again, includes the functional categories commonly used to explain (‘non-lexical’) syntactic phenomena. Her argument is thus a non sequitur.

It is certainly true that we do not have an account of variation in relative clause attachment, for a very good reason: at this point no one can say with any confidence what the underlying mechanisms of the variation are. Various accounts are on the market, some problematic for our approach and some not appearing to pose any problems of principle. The latter type are those that seek to explain the variation in terms of formal characteristics of the languages (e.g. Frazier and Clifton, 1996; Hemforth, Konieczny, Scheepers, and Strube, 1998). We have not developed an APT account, but there is no apparent reason to think that this cannot be done, once an adequate linguistic account is in place.

Compatibility issues

Our ‘mix and match’ approach naturally raises questions about the compatibility of the various elements we have incorporated. But so far at least we see no genuine cause for concern. To the extent that we do make what appear to be ‘hybrid’ proposals, we are trying to reinterpret conceptualizations from different approaches in terms that make them consistent within our own and, in most cases, Jackendoff’s model. Several of the commentators express concern about possible problems. But, interestingly, none make a case that any actually occur.

Carroll says we seem to need the levels of PF and LF (p. 23), apparently a corollary of her idea that we are dedicated Minimalists. We already offered reasons to believe that the absence of LF in our model causes no problems of any kind. Carroll does not address that discussion. Regarding PF, we have no idea why she thinks a Jackendovian PS will not work in our model, and we will not speculate on the matter.

Pienemann is concerned about our use of aspects of connectionism – i.e. activation levels gradually adjusted through experience – in conjunction with a largely symbolic approach. His argument is that Pinker and Prince (1988) showed that symbolic and connectionist approaches are irreconcilable and so our use of activation levels should be viewed with suspicion (p. 38). Our understanding of Pinker and Prince is quite different. They mentioned three possible views of connectionism, two of which are compatible with symbolic approaches. The third, ELIMINATIVE CONNECTIONISM, is not, but this hardly needed to be demonstrated, as the essence of this extreme variety is that rules are fictions and the only reality is the PDP network. This radical view obviously has nothing to do with our proposal, which maintains the rule-governed character of symbolic approaches and does not even include a connectionist network. So we do not understand Pienemann’s concern.

Dijkstra and Haverkort are uneasy about our use of activation levels with Chomskyan features. We would argue that a combination of this sort is not only feasible (no one has offered any reason to think it is not) but probably essential. Language, its acquisition, and its use are rule-governed. They are also fuzzy and non-discrete. Dealing with this dual nature of language is possibly the most important – and most challenging – problem facing the theorist. The use of features that have discrete values based on a continuum of possible strengths offers a promising approach.

More generally, if there is ever to be an adequate general theory, it will have to include reasonable, research-based accounts of (a) the nature of language, (b) the development of language, and (c) the use of language. In other words, it will have to find ways to incorporate the strengths of linguistics, cognitive psychology, and psycholinguistics. We believe we have at least taken a meaningful step in that direction.

What more needs to be done

As we said, our proposal is very ambitious, with the implication that LOTS still needs to be done. The commentaries provide many useful suggestions.

Dijkstra and Haverkort emphasize the importance of further developing working memory in the model, and we agree. Herschensohn focuses on the potential value of placing morphology at the center of our linguistic analyses, a suggestion we take very seriously. The possible connection with Dynamic Systems Theory (Dijkstra and Haverkort) is also quite interesting. Carroll points out some issues that are worth pursuing, such as relative clause attachment preferences and the use of gender markings in processing, though she chooses to present them as things APT cannot do. MacWhinney raises the interesting possibility of a more thorough comparison of French and

English adverb placement as a way to derive predictions about transfer. Roeper mentions the need to shed light on the syntax/semantics interface and the involvement of general cognition, an area where Jackendoff's approach seems particularly strong. Dijkstra and Haverkort point out the value of connecting APT to neuroscience research and of dealing with regular and irregular morphology. Topics that we find especially interesting for further development, and which came up in the commentaries, include cross-linguistic influence and the limited success of L2 learners relative to L1 learners. More generally, the approach clearly needs more explicit development in each of its areas, as we have always recognized.

That said, we feel that some of the commentators have greatly overstated their criticisms in this regard. When the initial presentation of a very broad approach is at issue, one can hardly go wrong in criticizing it for vagueness. But accusations of vagueness can easily become vague themselves, making the problems look far greater than they actually are. One example will suffice. Regarding the nature of items in the lexical stores, we started with Jackendoff's rather well-developed notions and went on to say a great deal about our own versions. In this context, Harrington's simple statement that 'what an item consists of is not specified' (p. 29) is quite misleading.

We recognize, once again, that we have not responded to all the issues brought up in the commentaries and may not have done justice to some of those we did discuss. We hope, in future, to be able to deal more comprehensively with the interesting questions raised by all our commentators.

References

Bates, E. & MacWhinney, B. (1982). Functionalist approaches to grammar. In E. Wanner & L. R. Gleitman (eds.), *Language acquisition: The state of the art*, pp. 173–218. Cambridge: Cambridge University Press.

Bates, E. & MacWhinney, B. (1987). Competition, variation, and language learning. In B. MacWhinney (ed.), *Mechanisms of language acquisition*, pp. 157–193. Hillsdale, NJ: Erlbaum.

Bates, E. & MacWhinney, B. (1989). Functionalism and the Competition Model. In B. MacWhinney & E. Bates (eds.), *The crosslinguistic study of sentence processing*, pp. 3–73. Cambridge: Cambridge University Press.

Bock, K. & Loebell, H. (1990). Framing sentences. *Cognition*, 35, 1–39.

Cuetos, F. & Mitchell, D. C. (1988). Cross-linguistic differences in parsing: Restrictions on the use of the Late Closure Strategy in Spanish. *Cognition*, 30, 73–105.

Edelman, G. M. (1992). *Bright air, brilliant fire: On the matter of mind*. New York: Basic Books.

Frazier, L. & Clifton, C. Jr. (1996). *Construal*. Cambridge, MA: MIT Press.

Guillelmon, D. & Grosjean, D. (2001). The gender marking effect in spoken word recognition: The case of bilinguals. *Memory and Cognition*, 29, 503–511.

Hemforth, B., Konieczny, L., Scheepers, C. & Strube, G. (1998). Syntactic ambiguity resolution in German. In D. Hillert (ed.), *Sentence processing: A crosslinguistic perspective*, pp. 293–312. San Diego, CA: Academic Press.

Jackendoff, R. (1997). *The architecture of the language faculty*. Cambridge, MA: MIT Press.

Jackendoff, R. (2002). *Foundations of language*. Oxford: Oxford University Press.

MacWhinney, B. (2001). The competition model: The input, the context, and the brain. In P. Robinson (ed.), *Cognition and second language instruction*, pp. 69–90. Cambridge: Cambridge University Press.

Pinker, S. & Prince, A. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. In S. Pinker & J. Mehler (eds.), *Connections and symbols*, pp. 73–193. Cambridge, MA: MIT Press.