Attention is a necessary construct for understanding virtually every aspect of second language acquisition. Both information processing and sociolinguistic accounts of variation assume that variation in use is a consequence of variation in attention. Attention is central to all accounts of the development of fluency. Understanding L2 development also invites the concept of attention. Even assuming a strong innateness position, at least the triggers of innate knowledge must be attended to, and in cognitive theories, attention to input plays an essential role in storage and hypothesis formation.

Attention also mediates between individual difference factors and SLA in at least three ways: attitudes and motivation make a difference because motivated learners attend more; one dimension of language aptitude is working memory, a construct which implies attention; and learning strategies are either strategies for focusing attention on language or for sustaining attention while doing something else in addition. Learner-external factors such as task requirements, task instructions, and all focus-on-form techniques (including explicit instruction) also affect what is attended to, thereby causing their effects.

The theoretical issues dealt with in this paper are organized around the basic assumptions, definitions, and metaphors concerning attention in the psychological literature and include the idea of attention as limited capacity, the notion of selective attention, and the role of attention in action control, access to awareness, and learning. Like most psychological concepts initially based on common experience, attention does not refer to a single mechanism, but to a variety of mechanisms with different functions. These include alertness, orientation, preconscious registration (detection without awareness), selection (detection with awareness within selective attention), facilitation, and inhibition. This does not diminish the centrality of attention (in its several manifestations) for learning. Although recent evidence using the negative priming paradigm indicates the possibility of some unattended learning, this appears limited in scope and relevance for SLA, and there is little doubt that attended learning is far superior. For all practical purposes, attention is necessary for all aspects of second language learning.

INTRODUCTION

The basic goal of this paper is to sketch—with necessarily broad strokes—part of a cognitive theory of second language acquisition that relies on the mental processes of
language learners as the basic explanation of learning.\(^1\) I am particularly concerned with those mental processes that are conscious, under the working hypothesis that SLA is largely driven by what learners pay attention to and notice in target language input and what they understand the significance of noticed input to be. This stands in opposition to what Jerome Bruner (1992) has called the "magical realist" view, that unconscious processes do everything.

A full understanding of the ways in which awareness shapes SLA is beyond the scope of this paper, however. Specifically, the issue of explicit and implicit learning (and related questions concerning the role of explicit and implicit knowledge) in SLA are not discussed here at any length. Both implicit and explicit learning surely exist, and they probably interact. Implicit learning (learning without awareness) is shown by numerous demonstrations that the result of allocating attention to stimulus arrays (input) results in more learning than can be reported verbally by learners. Knowledge of the grammar of one's first language is an obvious case. Native speakers of French "know" the rules for using the subjunctive, even if they know none of them explicitly. In experimental studies, it has also been shown that people can learn to control complex systems without recourse to an explicit mental model of how the system works (Berry, 1994). Various theories have been proposed to account for this basic phenomenon. In SLA, those most discussed at the present time are the UG account of things, which argues for unconscious deductive reasoning from innate principles, and the connectionist account, in which automatic, implicit learning results from the strengthening and inhibition of connections in an associative network, a simple, "dumb" process that leads to a complex and intelligent result (Elman, Bates, Johnson, Karmiloff-Smith, Parisi, & Plunkett, 1996). On both accounts, the learning is unconscious.

Explicit learning (learning with awareness) is also common. Probably most readers have learned a language recently enough to remember some of the experience or have learned some other cognitively demanding skill and can verify that learners commonly form (conscious) hypotheses about the target of their learning and modify those hypotheses as they encounter more information. What these two kinds of learning, implicit and explicit, have to do with each other continues to be a topic of great debate.

---

\(^1\) This paper is a revised version of a presentation at SLRF '97 (Michigan State University, October 17-19, 1997), part of a "point-counterpoint plenary" with Jacqueline Schachter of the University of Oregon. This paper, under the title "There is no learning without attention," was followed by one by Professor Schachter presenting the view that multiple types of evidence for unconscious learning of various kinds converge on the notion that unconscious adult learning can and does take place in some though presumably not all areas of language. A further revision of this paper is to appear in P. J. Robinson (Ed.), Cognition and second language instruction (Cambridge Applied Linguistics).
within SLA and elsewhere. In SLA, the question has frequently been posed in terms of whether or not "learned" knowledge can become "acquired" (Krashen, 1981) or whether the learner's conscious hypotheses can become internalized. Another, possibly more productive way to pose the question is in terms of learning processes (rather than types of knowledge), to ask whether bottom-up, data driven processing and top-down, conceptually driven processing guided by goals and expectations (including beliefs and expectations concerning the target language grammar, phonology, and so on) interact, to which the answer is probably yes, they do (Ellis, 1994a, 1996a, 1996b, forthcoming; Carr & Curran, 1994; Mathews, Buss, Stanley, Blanchard-Fields, Cho, & Druhan, 1989; Robinson, 1995).

The concerns dealt with in this paper are limited to issues related to the role of attention in SLA, and as Carr and Curran (1994) have pointed out, there are good reasons for separating the issues of attention and awareness in learning. Carr and Curran argue that focused attention is required for structural learning to take place, at least when complicated or ambiguous structures are the objects of learning, but allocating attention to a task does not guarantee conscious awareness of the structural learning that takes place. At the same time, it is very difficult to separate attention and awareness completely, because of the common assumption that if you are aware of something you are attending to it, and if you are attending to something you are aware of it (Carr & Curran, 1994, p. 219). The solution adopted to this problem here is to limit the discussion of attention and its subjective correlate, "noticing," to awareness at a low level of abstraction. "Noticing" is therefore used here in a restricted sense, as a technical term roughly equivalent to "clear perception" or to Tomlin and Villa's (1994) "detection within selective attention." My intention is to separate "noticing" from "metalinguistic awareness" by assuming that the objects of attention and noticing are elements of the surface structure of utterances in the input, instances of language, rather than any abstract rules or principles which such instances may be examples of. Although statements about learners "noticing [= becoming aware of] the structural regularities of a language" are perfectly fine in ordinary language, these imply metalinguistic reflection (thinking about what has been attended and noticed, forming hypotheses, and so forth), more than is implied by the restricted sense of noticing used here².

---

² As Truscott (forthcoming) has pointed out, for some in SLA rules are considered to be the targets of noticing (R. Ellis, 1993; Fotos, 1994).
ATTENTION IN CURRENT ACCOUNTS OF SLA

Even a cursory review of the SLA literature indicates that the construct of attention seems to be necessary for understanding virtually every aspect of second and foreign language learning.

Understanding Variation

Mellow (1996) has argued that when non-automatized knowledge is target-like but automatized knowledge is not, tasks for which attentional resources are abundant will result in more accurate language use than tasks for which attentional resources are limited. For example, redundant grammatical elements that have not been automatized are likely to be omitted in tasks that make high demands on attention such as comprehension tasks but will be supplied more consistently in tasks such as writing, which does not make as high demands on attention. Variability can also be induced by task constraints and instructions. Hulstijn and Hulstijn (1984) showed that performance on two Dutch word-order rules in a story retelling task improved when the subjects’ focus of attention was experimentally manipulated towards grammatical correctness. From a different perspective, Tarone (1996) has argued that language learners should not be viewed solely as decontextualized information processors, emphasizing that social context (including interactional pressures) is what causes a speaker to pay more or less attention to one or another linguistic form. However, the information-processing account and the social variationist account agree that variations in attention underlie variations in use.

Understanding Fluency

Attention is a key concept in all accounts of the development of fluency (Schmidt, 1992). Models that contrast controlled with automatic processing posit a transition from an early stage in which attention is necessary and a later stage (after practice) in which attentional resources are no longer needed and can be devoted to higher level goals (McLaughlin, Rossman, & McLeod, 1983; Shiffrin & Schneider, 1977). According to Logan’s instance theory (1988; Logan, Taylor, & Etherton, 1996), a competitor to the standard information processing view, the transition to fluency is not the result of developing automatic routines that do not require attention, but rather the replacement of slower algorithmic or rule-based procedures by faster memory-based processing. However, this theory is also based on some crucial assumptions about attention: encoding into memory is an obligatory consequence of attention (representations in memory are not
complete and accurate snapshots, but only encode what subjects pay attention to), and retrieval is an obligatory consequence of attention at the time of retrieval. Similarly, chunking theories of fluency assume a role for attention; chunking is a mechanism that applies automatically, but only to attended input (Servan-Schreiber & Anderson, 1986). Other models of fluency emphasize executive control and skilled selective attention. Bialystok has argued that the basis of fluency is the ability to focus attention on relevant parts of a problem to arrive at a solution (Bialystok, 1994; Bialystok & Mitterer, 1987).

**Understanding Development**

Understanding L2 development also invites the concept of attention, although some accounts emphasize attention more than others. If one is concerned only with linguistic competence and subscribes to a strong innateness position, that development is the mere triggering of innate knowledge, presumably at least the triggers have to be attended to (Schmidt, 1990). Within the connectionist position (which does not distinguish between competence and performance), input and output units are assumed to be attended, although many accounts are silent on this issue (for one that is explicit, see Cohen, Dunbar, & McClelland, 1990).

The role of attention is emphasized most in cognitive approaches to language development, such as those of Pienemann (1983) and Wolfe-Quintero (1992), within which attention to input is essential for storage and a necessary precursor to hypothesis formation and testing (see also Slobin 1973, 1985). Common to these approaches is the idea that L2 learners process target language input in ways that are determined by general cognitive factors including perceptual salience and the continuity of elements (Towell & Hawkins, 1994).

Peters (1998) proposes that in every domain of language learning (phonology, grammar, semantics, pragmatics, vocabulary, discourse structuring), learners must attend to and notice any source of variation that matters, whatever makes a difference in meaning. For example, in syntax, one may say in English both “I turned the covers down” and “I turned down the covers,” but there is no difference in meaning that depends on the position of the direct object. Native speakers do not attend to this difference, and non-native speakers do not have to attend to it either, at least for comprehension (to produce both variants, it is presumably necessary for them to notice each possibility, unless both are assumed possible based on L1 knowledge). However, if an utterance contains a pronoun, then there is a difference: “I turned it down” is possible, but “I turned down it” is possible only in the sense of “I turned down the road,” while “I turned the road down” makes sense only with the semantic reading of a road being offered but
rejected as a gift. In this case, Peters argues that learners do have to notice the difference in ordering and be aware that it matters (mapping forms with their appropriate meanings). Moreover, since beginning learners are cognitively overloaded, they cannot pay attention to all meaningful differences at once. If they have not learned what is simple, they cannot learn what is complex, but as simpler processing routines are over-learned, they have more capacity to attend to details, eventually being able to attend to whatever native speakers pay attention to. In the multidimensional model of Pienemann and Johnston (1987), developmental features and natural orders are related to the learner’s processing space and the freeing of attentional capacity. For example, the crucial point for accurate production of third person -s is that the learner must have enough processing space available to generate a third person marker and keep it in working memory until the appropriate moment arrives for attaching it to a verb.

In SLA, the proposal that attention is both necessary and sufficient for learning L2 structure has been well expressed by VanPatten (1994):

Bob Smith is a learner of Spanish, a language that actively distinguishes between subjunctive and indicative mood ... He begins to notice subjunctive forms in others’ speech. He attends to it. Soon, he begins to use it in his own speech, perhaps in reduced contexts, but nonetheless he is beginning to use it. If you ask him for a rule, he might make one up. But in actuality, he doesn’t have a rule. All he knows is that he has begun to attend to the subjunctive and the context in which it occurs and it has somehow begun to enter his linguistic system... Bob did not need to come up with a conscious rule; he only needed to pay attention. (p. 34)³

Understanding Individual Differences

Attention is a useful construct for understanding individual differences in SLA. As Tremblay and Gardner (1996) have pointed out with respect to motivation, a statement that some aspect of motivation leads to higher proficiency or better performance does not answer the question of why such a relationship exists. Models of motivation and learning can be improved by the identification of mediators that explain why one variable has an

³ While VanPatten’s account is generally in accord with the argument presented in this paper, it could be made more precise. It is not the “subjunctive” itself that needs to be paid attention to, but instances or exemplars of it. Clearly it is not necessary that learners know the label “subjunctive” or that they have a metalinguistic rule (though this may be helpful). However, if Peters (1998) is correct in arguing that learners must notice variation that makes a difference, then noticing individual instances should lead to success in using formulaic utterances with subjunctive forms but fully productive use (when a given verb can appear in a particular context in either subjunctive or indicative form) would require that learners attend to the mapping between forms and meanings.
effect on another. In a revised version of Gardner’s well-known socio-educational model, Tremblay and Gardner propose that three motivational behaviors—effort, persistence, and attention—are the mediators between distant factors such as language attitudes, motivational factors (value, self-efficacy, and goal-salience), and achievement, finding support for a LISREL structural equation model linking these variables in a study of achievement in French courses. In addition, three studies to date (MacIntyre & Noels, 1996; Oxford & Nyikos, 1989; Schmidt, Jacques, Kassabgy, & Boraie, 1997) have found strong links between motivation and learning strategies, particularly cognitive and metacognitive strategies. These strategies are either strategies for focusing attention on some aspect of the target language or for sustaining attention while doing something in addition—inferencing, looking for patterns, monitoring (paying attention to one’s output and to the process of learning itself), and other types of active conscious processing (O’Malley, Chamot, & Walker, 1987; Oxford, 1990).

At least one aptitude factor, working memory capacity (Baddeley, 1986; Ellis, 1996a; Harrington & Sawyer, 1992), is closely related to attention. Working memory capacity refers to the ability to rehearse and process sequential information (for example, repeating strings of digits or repeating a set of sentences while simultaneously remembering the final word of each sentence). Robinson (1995) has suggested that my concept of “noticing” can be redefined as detection plus rehearsal in working memory. Baddeley, Papagno, and Vallar (1988) have argued that such rehearsal is necessary for learning unfamiliar verbal material, although not necessary for forming associations between meaningful items that are already known.

**Understanding the Role of Instruction**

Sharwood Smith (1995) points out that input salience can be internally derived (input becomes noticeable to the learner because of internal cognitive changes and processes) or externally derived (input becomes more noticeable because the manner of exposure is changed). One likely role of explicit instruction is that by changing expectations, it helps focus attention on forms and meanings in the input, a prerequisite for subsequent processing (de Graaff, 1997; N. Ellis, 1993; R. Ellis, 1994; Hulstijn & DeGraaf, 1994; Long, 1988; Schmidt, 1990; Schmidt & Frota, 1986; Sharwood Smith, 1993, 1994; Tomlin & Villa, 1994; Terrell, 1991; VanPatten, 1994). It can be argued that task requirements, task instructions, and input enhancement techniques affect what is attended to and noticed in on-line processing, thereby causing their effects (Doughty, 1991; Doughty & Williams, in press; Skehan, 1996).
Similar characteristics of informal instruction, ranging from immersion contexts to natural interaction with native speakers of a language, have also been widely commented upon (Pica, 1994, 1997). Long (1983, 1992, 1996) has argued that interactional modifications such as clarification requests and recasts, are more consistently present than are input modifications (e.g., linguistic simplification) in interaction between native and nonnative speakers and that the nature of interactional modifications as attention-focusing devices is what makes them likely to be helpful for acquisition. Gass and Varonis (1994) have proposed that interaction serves to focus learners' attention on form in instances where there is perceived difficulty in communicating, "raising to awareness that area of a learner's grammar that deviates (either productively or receptively) from native speaker usage." Swain (1985, 1993; Swain & Lapkin, 1995) has proposed that one reason learners in immersion contexts exhibit weaknesses in grammatical accuracy even after receiving years of comprehensible input is that they are not called upon to produce much, arguing that "producing the target language may be the trigger that forces the learner to pay attention to the means of expression needed in order to successfully convey his or her own intent" (1985, p. 249).

If all these accounts are correct, attention is a crucial construct for SLA. Attention is the pivot point at which learner-internal factors (including aptitude, motivation, current L2 knowledge, and processing ability) and learner-external factors (including the complexity of input, discoursal and interactional context, instructional treatment, and task characteristics) come together. What then happens within attentional space determines language development, including the growth of knowledge (establishment of new representations), fluency (access to that knowledge), and variation.

However, it could be argued that attention in these accounts is merely a deus ex machina that does not actually explain anything. At the least, one must wonder whether a unitary concept of attention based on ordinary experience or folk psychology can be the explanation of so many varied phenomena. To gain a better understanding of what attention is and how it works, it is necessary to turn to psychology, where attention has been a major focus of theory and empirical research for over a century, and to examine some of the assumptions, definitions, metaphors, theoretical disputes, and empirical findings from that field.

**ATTENTION IN PSYCHOLOGY: SOME BASIC ASSUMPTIONS**

In psychology, the basic assumptions concerning attention have been that it is limited, that it is selective, that it is partially subject to voluntary control, that attention controls
access to consciousness, and that attention is essential for action control and for learning. All of these basic points were raised early on. The classic work on attention is that of William James (1890), who noted that “Everyone knows what attention is. It is the taking possession by the mind, in clear and vivid form, of one out of what seem several simultaneously possible objects or trains of thought. Focalization, concentration of consciousness, are of its essence. It implies withdrawal from some things in order to deal more effectively with others” (p. 403). The nature and mechanisms of attention turned out not to be so self-evident, however, and the topic continues to be discussed within an enormous literature, most of which is experimentally based.

Attention is Limited

The classic view in psychology is that limited capacity is the primary characteristic of attention (Broadbent, 1958; Kahneman, 1973), and this view has been taken on by many in SLA (McLaughlin, Rossman, & McLeod, 1983; VanPatten, 1994). Within this general view, some have stressed that there are two general human information processing systems. Such accounts contrast effortful, attention-demanding (“controlled”) processes with capacity-free (“automatic”) processes (Shiffrin & Schneider, 1977). Another variant of the basic notion of capacity limitations in attention is that of Wickens (1980, 1984), who proposed multiple, specific resource pools for processing stages, brain hemispheres, and modalities (visual, auditory, vocal, manual). This model accounts for the fact that attention-demanding activities can be carried out at the same time more easily if they call upon different modalities than if they draw upon the same modality. In other words, there is some flexibility to capacity limitations, though each resource pool is assumed to have limited capacity.

Since the 1960’s, when computers with limited memory systems began to come into widespread use, the primary metaphor for the limited capacity notion of attention has been memory (CPU). Another influential variant of the limited capacity metaphor for attention was influenced by psychoanalysis; Kihlstrom (1984) proposed that “effort” (analogous to Freud’s libido) was the limited resource.

Attention is Selective

In the classic account of attention, the common view is that a second characteristic of attention, that it is selective, is a corollary of limited capacity. Because there is a limited supply of attention and because any activity that draws upon it will interfere with other activities requiring it, attention must be strategically allocated. The basic metaphor here is economic (Shaw & Shaw, 1978). When resources are limited, a cost-benefit analysis
determines the focus of attention. VanPatten has drawn upon this metaphor in SLA, arguing that what is important in most SLA contexts is the meaning of messages, so limited attentional resources are directed first as those elements that carry message meaning (primarily lexicon) and only later (when the cost comes down) towards communicatively redundant formal features of language (VanPatten, 1990, 1994; Lee, Cadierno, Glass, & VanPatten, 1997).

In the selective attention literature, the most enduring controversy has been whether selection happens early or late in processing. One influential early view held that attending to one message eliminated perception of another (Broadbent, 1958). The metaphor here is that of a filter, gate, or bottle-neck. Later findings showing that individuals process highly meaningful words outside an attended channel (for example, in dichotic listening studies, in which different messages are played to the two ears) led some researchers to make the strong assumption that all information in the input stream is perceptually processed and that selection happens late (Deutsch & Deutsch, 1963).

Whether or not early selection occurs is still controversial (LaBerge, 1995), but more recent work in psychology has moved away from this issue and from the notion of selection as a corollary of limited capacity, identifying selection itself as the basic function of attention and emphasizing that selection may have other functions in addition to the allocation of scarce resources. Treisman (1992) argues that visual attention serves to integrate the features that belong to the same visual object, that is, to coalesce the properties of an object into a coherent perceptual representation (see also Shapiro, Arnell, and Raymond, 1997). Within the language acquisition field, Bialystok (1994) has also emphasized the importance of selection ("control" in her framework) rather than limited capacity as the primary characteristic of attention.

**Attention is Subject to Voluntary Control**

LaBerge (1995) emphasizes the importance of preparatory attention, which includes such things as being ready to step on the gas when a traffic light turns green or waiting to applaud at the exact moment the last sound of a musical performance ends. The benefits of preparatory attention include accuracy in perceptual judgment and categorization, as well as accuracy and speed in planning and performing actions. More generally, we can choose to pay attention to one stimulus (or some feature of a stimulus) over another, and a great deal of language teaching practice is founded on the premise that learners can attend to different aspects of the target language and that one of the important functions of teaching is to help focus learners’ attention. Hulstijn and Hulstijn (1984) have suggested that certain tasks can be repeated, with the teacher telling them each time to pay attention
to different features, such as grammar, pronunciation, rate of speech, completeness of information, and so on.

The idea that we can control the focus of attention is an ancient one, often equated with the soul or will. In the 19th century, Wundt and James were the most prominent proponents of this view of attention (Neumann, 1996), while recognizing as well that there is also a passive, involuntary form of attention. For example, one attends to a loud noise, whether one wants to or not. In the well-known Stroop effect, the printed name of a color word like “red” or “green” tends to interfere with the ability to name the color ink in which the word is printed, e.g., “brown,” when the word “red” is printed in brown ink (Dalrymple-Alford & Budayr, 1966). Involuntary attention is data driven, elicited bottom-up. Voluntary attention is top-down in the sense that attention is directed to outside events by inner intentions.

Attention Controls Access to Consciousness

To quote William James again, “My experience is what I agree to attend to” (1890, p. 403). The idea that one of the roles of attention (whether voluntary or involuntary) is to control access to consciousness is an old one (Baars, 1988, pp. 301-324; Shapiro, Arnell, & Raymond, 1997). Ever since Aristotle described the phenomenon of selective attention, consciousness (awareness) has been equated with the phenomena of limited capacity and selective access. Descartes assumed that selectivity occurred at a specific place, the pineal gland, between the mechanical brain processes shared with all animals and higher mental processes that are unique to humans. The idea that attention and awareness are essentially two sides of the same coin played a prominent role in most 19th century theorizing in psychology. For Wundt, the focus of consciousness was determined by the direction of attention, what he called apperception. Theodore Ziehen argued against Wundt’s concept of apperception, but also identified attention with access to awareness. In Ziehen’s model, attentional selection was based on competition. Sensations and latent (nonconscious) ideas compete for access to consciousness (a very contemporary view). Only the most strongly activated ideas and their associated sensations actually enter consciousness; those that do not remain unconscious and have no effect at all. A modern version of these ideas is that of Marcel (1983), who identifies focal attention as the mechanism that controls access to awareness and establishes the boundary between an early processing stage that produces nonconscious representations of all stimuli in a purely bottom-up manner and a higher state of phenomenal experience which consists of the imposition of a particular interpretation, a constructive act. Neisser’s (1967) model of attention was similar, in the sense that he viewed the essential
function of focal attention in terms of a constructive, synthetic activity (largely driven by top-down processes such as expectations) that makes stimuli available for further analysis. (See Neumann, 1996, for an excellent historical account of these views.)

Similar views are prominent in the major research paradigms that have been used in psychology to investigate attention. In reviewing two independent research traditions, one rooted in filter theory and largely investigated through dichotic listening studies, the other based on paradigms from visual information processing, Neumann (1990, 1996) identified the following assumptions in common:

1. Selection is the mechanism that moves information from one stage of processing to a subsequent stage;
2. The locus of selection is situated between the unlimited-capacity and the limited-capacity portions of the information processing system;
3. Selected stimuli are represented in conscious awareness, and unselected stimuli are not so represented.

The predominant metaphor of this view of selection as access to consciousness is that of the spotlight or zoom lens (Baars, 1996; Eriksen & St. James, 1986; Posner & Peterson, 1990). Various SLA theorists have also stressed the role of attention as the process that brings things into awareness. Bialystok has proposed a model of the development of L2 proficiency built around two cognitive processing components, called analysis and control. Analysis is the process by which internal, formal representations are constructed. Relatively unanalyzed representations (e.g., formulaic chunks useful for conversational purposes) gradually evolve into more analyzed representations which are required to support higher literacy skills. Control refers to access to these representations, the basis of fluency. For Bialystok (1994), awareness (consciousness) is the result of an interaction between analysis and control; the process of focusing attention onto internal representations “gives rise to the subjective feeling of awareness that has been called consciousness” (p. 165).

**Attention is Essential for Action Control**

The essential contrast here is between novice behavior, for which controlled processing is required, and expert behavior, which can be carried out either automatically (without attention) or mostly automatically with only a controlled assist at critical decision points (Reason, 1984). Schneider and Detweiler (1988) have proposed a model in which automatization is viewed as a gradual, continuous transition through five identifiable phrases: fully controlled processing, context-maintained controlled
comparison, goal-state-maintained controlled comparison, controlled assist of automatic processing, and fully automatic processing.

Attention is Essential for Learning

The orthodox position in psychology is that there is little if any learning without attention (Carlson & Dulany, 1985; Fisk & Schneider, 1984; Kihlstrom, 1984; Logan, 1988; Nissen & Bullemer, 1987; Posner, 1992; Shiffrin & Schneider, 1977; Velmans, 1991). This claim is often related to models of memory. It is argued that unattended stimuli persist in immediate short-term memory for only a few seconds at best, and attention is the necessary and sufficient condition for long-term memory storage to occur. In SLA as well, the claim has been made frequently that attention to input is necessary for input to become intake that is available for further mental processing (Carr & Curran, 1994; Gass, 1997; van Lier, 1991, 1994; VanPatten, 1994; Schmidt, 1995).

FROM METAPHORS TO MECHANISMS

Metaphors (attention as economic resource, pipeline, gate, executive, or spotlight) are useful and often drive research programs, but are pre-theoretic, whereas mechanisms seek to be explanatory. In a widely read article in SLA, Tomlin and Villa (1994) have drawn upon the work of Posner (1994; Posner & Peterson, 1990) to identify three mechanisms or subsystems of attention, each with identifiable neurological correlates: alertness, orientation, and detection.

Alertness vs. Orientation

In Posner’s account of attention, orienting refers to committing attentional resources to sensory stimuli. In earlier accounts (Pavlov, 1927), orienting was traditionally studied as a reflexive response (e.g., attending involuntarily to a loud noise), and Posner comments that the orienting system responds to involuntarily processed stimuli during early visual processing, but for Posner the essential characteristic of orienting is the alignment process itself (attributed to the posterior cortex), not whether such alignment is voluntary or involuntary. Orientation can therefore be modulated by a second attentional subsystem, the alertness or executive attentional system (in the mid-frontal lobe), which maintains a state of vigilance to increase the rate at which high priority information can be detected in the service of current goals.

Tomlin and Villa argue that both orientation and alertness are important in SLA. Alertness is related to motivation, interest in the L2, and classroom readiness to learn.
Orientation is related to such instructional techniques as input-flooding and VanPatten’s notion of attention to form: “that is, the learner may bias attentional resources to linguistic form, increasing the likelihood of detecting formal distinctions but perhaps at the cost of failing to detect other components of input utterances” (Tomlin & Villa, 1994, p. 199).

**Detection: Non-conscious Registration vs. Conscious Perception**

For Tomlin and Villa, both orientation and alertness enhance the likelihood of detection, but it is detection itself, the cognitive registration of stimuli (mapped to an anterior attention network that includes the anterior cingulate gyrus and nearby motor areas), that is the necessary and sufficient condition for further processing and learning. Tomlin and Villa stress that, in their view, detection is not equivalent to awareness. It is therefore necessary to distinguish between detection without awareness (for which “registration” seems a non-controversial choice of labels) and detection within focal attention accompanied by awareness (conscious perception or noticing).

Making this distinction, unfortunately, immediately raises some thorny methodological issues:

1. How can we know whether some stimulus (or a feature of it) has been attended?
2. How can we know whether some stimulus (or a feature of it) has been consciously noticed?
3. How can we know whether some stimulus (or a feature of it) has been registered, even if not attended or noticed?

Evidence that some part of target language input has or has not been attended to and noticed is sometimes available from learner productions (in either naturalistic or classroom settings), as in the following exchange reported by VanPatten (1990):

Interviewer: Cómo están ellos? (How are they?)
Subject: Son contento. (They are happy.)

Interviewer: Y ellos, cómo están? (And them, how are they?)
Subject: Son contento también. (They are happy too.)

VanPatten argues that in the above example the learner was so intent on meaning, that features in the input (verb choice between *ser* and *estar*) that were not crucial to meaning were neither attended nor noticed.

Learner reports, for example in diary studies, provide another source of information about what learners pay attention to and notice. Schmidt and Frota (1986) reported the results of a diary study in which there were so many instances of L2 use matching the learner’s reports of what had been noticed in input (in interaction with native speakers)
that the study supported the hypothesis that there is no language learning without attention and noticing. On the other hand, Warden, Lapkin, Swain, and Hart (1995) found no particular relationship between the quantity and quality of linguistic observations recorded by high school students of French in their journals during a summer exchange program and their progress as reflected in test scores. This might be because some learners found their language observations more worthy of report than did others, but as Tomlin and Villa point out, the essential weakness of diary studies is that their temporal granularity is too coarse: “Diary studies encompass spans of time as large as several weeks, but the cognitive processing of L2 input takes place in relatively brief spans of time, seconds or even parts of seconds” (1994, p. 185).

Leow (1997) has used data from think-aloud protocols produced by learners of Spanish completing a problem-solving task (an L2 crossword puzzle) to gather finer-grained data and to distinguish between two levels of awareness illustrated below by two responses to the same item (the stem changing verb mintieron): simple noticing (registration with awareness indicated by a report or repetition) and noticing with metalinguistic awareness:

Simple noticing:
“...the opposite of no, so it is supposed to be si, so 11 across is gonna be
mintieron (fills in mintieron)”

Noticing with metalinguistic awareness:
“... now let's see where is number 17 down? oh se durmieron, con a with a
u...repetir, ellos repetieron? I think it has a stem change, 25 down ir,
yes!...four down would be tu, so durmió (writes in durmió) done cool, I like
this. Number 5 ellos of pedir, that asked, pidieron and it's good ... nine is
gonna be si again mintieron and obviously I spelled number 11 wrong so I can
fix that (changes mentieron to mintieron)”

This technique appears to have been successful in distinguishing between two levels of awareness (those who showed higher levels of awareness learned more than those whose protocols showed that they merely attended and noticed), but it is difficult to see how such techniques could show that subjects did not attend or notice something, since verbal reports (even when concurrent) cannot be assumed to include everything that is noticed. Jourdenais, Ota, Stauffer, Boyson, and Doughty (1995) have also used think-aloud protocols to see whether learners in a focus-on-form instructional treatment (target items were textually enhanced) would notice and learn more than subjects in a comparison group. Enhancement participants did notice more target forms and did produce more
target features in production, but once again the methodology cannot identify examples of target features that were not noticed but might have been non-consciously registered.

In computer-assisted learning contexts, instructional treatments can be designed to focus learners’ attention on crucial aspects of input, and the success of such efforts can be assessed not only through learners’ reports of what they notice (Chapelle, 1998) but also with programs that track the interface between user and program, for example by recording mouse clicks and eye movements (Crosby, 1998). However, these records can provide information only about orientation, while detection remains invisible. Nevertheless, it is possible to operationalize the distinction between non-conscious registration and conscious noticing within focal attention in some experimental settings. Merikle and Cheesman (1987) have introduced a distinction between the objective and subjective thresholds of perception. The clearest evidence that something has been consciously perceived or noticed (exceeded the subjective threshold) is a verbal report of awareness or recall. However, many have argued that it is unreasonable to assume that the absence of a verbal report means that something was not noticed, since verbal reporting requires reflexive self-description and adequate language to describe what is noticed. One might reasonably claim to have noticed or consciously perceived a particular feature of a regional accent of English, without being able to describe it accurately or to reproduce it very accurately. A less demanding but more reasonable criterion for awareness is above chance performance in a forced-choice recognition test. For example, if subjects are presented with a series of utterances in Spanish and immediately afterwards forced to identify the verb form, preterit or imperfective, that occurred in each utterance and can do so, that is reasonable evidence that they noticed them, at least in cases where it is possible to be sure that they were not generated from internal knowledge. More importantly, if subjects cannot identify which forms occurred in input with better than chance accuracy, that is much stronger evidence for the absence of noticing than their inability to produce them. If it can be shown (using more indirect measures) that these subjects did, nevertheless, register or detect the forms that they cannot identify in a recognition test, then we would have a strong case for pre-attentive, unconscious registration.

One widely accepted indirect measure of registration is priming, the facilitation of responses to one stimulus by the prior presentation of a related stimulus. Marcel (1983) showed that subliminally presented words that subjects could not consciously see could prime semantic associates. Eich (1984) reported dichotic listening experiments in which pairs of words were both presented to the unattended ear, one of which was ambiguous (e.g., *fair* or *fare*) while the other biased its less common interpretation (e.g., *taxi*).
Recognition of both members of the pair was poor (indicating that the subjective threshold of perception was not exceeded), but in a spelling test subjects were biased in the direction of the disambiguated meaning (indicating that the objective threshold of perception had been reached). However, these studies used well-known native language words, leaving unanswered the question of whether novel material (such as foreign language input) can be unconsciously registered in a similar fashion. Before turning to the evidence for this possibility, it is necessary to discuss two additional mechanisms of attention.

**Facilitation vs. Inhibition**

The theories discussed so far have assumed that attention directly facilitates or enhances processing (LaBerge, 1995). Automatic processing is assumed up to some level. Attention then intervenes, enabling selected information to receive further processing (Neill, Valdes, & Terry, 1995), and ignored stimuli are not processed further (Van der Heijden, 1981). However, there is a logical alternative. The mechanism of attention might instead inhibit, block, or suppress the processing of irrelevant information, so that processing of relevant information simply proceeds without interference from irrelevant information.

The inhibition construct played an important role in early psychological theories (Harnishfeger, 1995). Luria (1961) for example, demonstrated a developmental sequence of action control via verbal regulation. Very young children are not able to guide their own behavior by inhibiting irrelevant behavior either through external or internal speech. Later, toddlers become capable of using external verbal commands to direct their behavior, yet they continue to be unable to regulate their behavior with their own verbal instructions (internal or aloud). However, until recently most psychological models paid little attention to inhibition. Inhibition was totally ignored by behaviorists, and during the period in which cognitive psychology was dominated by information-processing models, inhibitory mechanisms were not seen as particularly useful (Harnishfeger, 1995)

Selective attention probably cannot be only an inhibition mechanism, because outside of laboratory experiments there are just too many things that would have to be inhibited. However, resistance to interference from potentially attention-capturing processes is clearly important in understanding many real life phenomena, such as school children who cannot keep their attention on class, adults with obsessive-compulsive disorder (who cannot inhibit the intrusion of unwelcome thoughts), and older adults with Alzheimer’s disease (Kane, Hasher, Stolzfus, Zacks, & Connelly, 1994). Considerable evidence has accumulated in support of the suppression position in general (Tipper, 1985; Tipper &
Baylis, 1987; Tipper & Cranston, 1985), and research on inhibitory processes is probably the most active and theoretically interesting work within attention theory at the present time. The reasons for this probably include the growing preference among cognitive psychologists for the brain metaphor rather than the computer metaphor, the impact of studies relating the frontal cortex to resistance and interference, and connectionist models in which simple processing units send both excitatory and inhibitory signals to each other (Bjorklund & Harnishfeger, 1995; Dempster, 1995).

Evidence for facilitation and inhibition effects comes from positive and negative priming phenomena, respectively. Suppose that a subject must attend and respond to one source of information, ignoring another and, subsequently, must unexpectedly respond to the just-ignored information or to entirely new information. According to facilitation theories, the effects of the ignored object dissipate over time, but if those effects have not dissipated completely, then processing of the previously ignored object should show an advantage over completely new information (positive priming). If, however, if the processing of ignored information is inhibited when processing attended information, subsequent processing of the ignored information must overcome any persistence of that inhibition (negative priming). Ignored (inhibited) information should be more difficult to process than new information (Neill, Valdes, & Terry, 1995).

Negative priming has been demonstrated in a great variety of selective-attention tasks. For example, research on lexical ambiguity has focused heavily on the question of whether sentence context is able to constrain processing only to the meaning that is appropriate in the context, or whether multiple meanings become activated in all contexts. Data supporting the latter view (temporary non-conscious activation of multiple meanings) are frequently cited as evidence for the modularity of lexical processing. However, Simpson and Kang (1994) and Yee (1991) report recent studies that are concerned with the fate of meanings after the processing of an ambiguous word has presumably run its course, reporting several studies showing that one meaning of an ambiguous word is suppressed following the selection of the other for a response.

**CAN THERE BE LEARNING WITHOUT ATTENTION?**

It is necessary to deconstruct this question somewhat in order to find the essential issues, since different variants of the question may well have different answers. For example, one might paraphrase the question as “Is it necessary to pay attention, deliberately, to some aspect of second language input in order to learn it?,” with a focus on intentionality, alertness, and voluntary orienting to specific stimuli. Because we know
that attention can be involuntarily attracted to stimuli or stimulus features (involuntary orienting), it cannot be claimed that learners must *intentionally* focus their attention on each particular aspect of L2 input in order to learn it. Even if it is true that in order to learn anything one must attend to it, that does not entail that it is necessary to have either the intention to attend or the intention to learn. On the other hand, since we know that preparatory attention and voluntary orienting vastly improve encoding (LaBerge, 1995; Cowan, 1995) and since many features of L2 input are likely to be infrequent, non-salient, and communicatively redundant, intentionally focused attention may be a practical (though not theoretical) necessity for successful language learning. Language learners who take a totally passive approach to learning, waiting patiently and depending on involuntary attentional processes to trigger automatic noticing, are likely to be slow and unsuccessful learners.

Another way to paraphrase the question is to ask whether all aspects of L2 input must be attended in order to learn them, or whether some kind of global attention to input is sufficient. For example, is it sufficient for attention to be focused on meaning, with message form picked up without any attention to it (Paradis, 1994), or is Gass (1997) correct in arguing that apperceived input that is processed only semantically (for example, with the help of non-linguistic cues, isolated lexical items, and contextual expectations) and receives no syntactic processing will not lead to development of syntax? Is it the case that nothing is free, that in order to acquire phonology, one must attend to the sounds of target language input; in order to acquire pragmatics, one must attend to both the linguistic form of utterances and the relevant contextual features; in order to acquire L2 syntax one must attend to the order of words and the meanings they are associated with (Schmidt, 1990, 1993b, 1995)? It has been suggested that aspects of language may differ in their attentional requirements; perhaps learning lexicon and morphology require attention in ways that learning syntax does not (VanPatten, 1994; Schwartz, 1993). Based on a review of vocabulary learning studies with amnesics, Ellis (1994b) concludes that attention (but not awareness) is necessary and sufficient for learning the perceptual aspects of novel word form, while learning word meanings requires both attention and explicit awareness.

The psychological literature provides less help in resolving this question than some others concerning attention, because relatively few studies have assessed the effects of focusing attention on different features of stimulus sets (as opposed to dividing or sharing

---

4 We also know that learners are not free to allocate their attention wherever they wish. VanPatten (1990) has carried out experiments showing that learners have great difficulty in attending to both form and content simultaneously, although they need to do both in order to map form and meaning, the essential task of language learning.
attention between two sources of stimuli). The few studies that address this issue suggest that stimulus attributes are filtered by attention and only those that are relevant to the experimental task and receive attention are represented in stored instances (Logan, Taylor, & Etherton, 1996). This is sufficient grounds to motivate an attentionally determined encoding specificity hypothesis for SLA, but insufficient to settle the issue, which clearly requires research within SLA itself, focused on different domains of language.

The question of whether global attention to L2 input might be sufficient is also reflected in the sentiment expressed by many SLA researchers that many features of language could not possibly be attended to, because they are too subtle and abstract (Sharwood Smith, Internet posting to PSYCHE-B@listserv.uh.edu, 11/3/97). If these features of language are taken to be those that modern linguistic theory makes explicit—such as abstract principles of government, constraints on movement, and the like—then this must surely be true, since many of these are unrecognizable in any conscious way simply by attending to input. On the other hand, assuming that abstract grammars of the type described by linguists are what L2 learners acquire (itself not an uncontroversial assumption), it may well be necessary for learners to attend to the evidence for these principles, including the presence of morphological material, the order of elements, and the ways in which specific utterances map onto meanings. That is, the issue may be resolved in terms of the distinction made at the beginning of this paper between attention to utterances as distinct from an understanding of abstract rules or principles.

Another way to resolve this issue is suggested by Sharwood Smith's (1994) distinction between competence, held to be elusive, intuitive, and essentially subconscious, and the on-line production and reception processes of actual language performance. Sharwood Smith gives as an example the position of a verb in a sentence. As far as competence is concerned, there is no rule as such for positioning a verb. Its position falls out from some general principles of universal grammar which prevent it from going anywhere but its one position. However, language processing takes place in real time, so a verb (whose position in the structure of a sentence may not be governed by a rule) must be positioned appropriately (by a rule or routine) in real time utterances generated by a language processor. In order to establish such routines, attention must be allocated to the order of elements (sequences) in both input-processing and in production.

Probably the most interesting variant of the question of whether attention is necessary for learning in all cases concerns detection in Tomlin and Villa's sense and the distinction that can be made between nonconscious registration and conscious perception, "noticing," or—in Tomlin and Villa's terms "detection within selective attention."
Tomlin and Villa’s claim that detection is necessary for learning but that detection need not result in awareness is somewhat anticlimactic, since it is self-evidently true that some aspect of language that is not registered in any sense will not lead to learning. The most interesting question, and the hardest to answer conclusively, is whether selection accompanied by awareness is necessary, or whether pre-consciously detected information is sufficient for learning.

There is evidence for the cognitive registration of stimuli without focal attention or awareness, both from subliminal perception studies and from studies using measures of implicit memory to establish the registration of unattended information (Schmidt, 1990, 1993a, 1994a, 1994b, 1995). These studies clearly show cognitive activation (for about a tenth of a second) of previously well-learned information present in long term memory. However, the vast majority of these studies do not show learning of anything new. On the basis of this distinction, I have proposed a strong version of the “noticing hypothesis,” a claim that while there is subliminal perception, there is no subliminal learning.

Several types of studies have the potential to falsify this claim, but each entails methodological difficulties. The most straight-forward would be to demonstrate the existence of subliminal learning directly, either by showing positive priming of unattended and unnoticed novel stimuli or by showing learning in dual task studies in which central processing capacity is exhausted by a primary task. The problem with positive priming studies is that one can never really be sure that subjects did not have at least fleeting awareness of what they could not report (DeSchepper & Treisman, 1996; Merkle & Daneman, 1995). The problem with dual task experiments is that one cannot be sure that no attention is devoted to the secondary task, and in experiments using this paradigm, selective attention procedures vary considerably in the extent to which they permit or even prompt subjects to divide attention between two information sources rather than focusing exclusively on one (Greenwald, 1992). Schachter, Rounds, Wright, and Smith (1996) have reported learning of complex WH-questions in a “nonattentitional” condition. However, subjects in this condition were required to read the target structure sentences out loud while performing on-line a substitution of an earlier seen word for its synonym in the target sentence. Since reading aloud requires attention to the sequence of words and Schachter et al. define the learning task as one of serial learning, it is very difficult to accept the claim that this was truly a nonattentional condition. If not—and both Schachter et al. and Curran and Keele (1993) comment that their use of “nonattentional” does not mean that no attention at all is devoted towards the secondary task—then the results are compatible with the idea that performance under conditions of divided attention results in some learning (though less than in single task learning) as
long as the competing task does not complete deplete attentional resources. The general point is that both types of demonstration, positive priming and dual task learning, are likely to be contaminated by conscious processes.

Jacoby, Lindsay, and Toth (1992) have argued that the trick in demonstrating nonattentional learning is to use the logic of opposition, to arrange experiments in which unconscious processes oppose the aims of conscious processes. DeSchepper and Treisman (1996) have recently used this logic to produce what may be the most solid demonstration to date of unattended learning, using an experimental paradigm devised by Rock and Gutman (1981). Rock and Gutman presented two overlapped nonsense shapes and asked subjects to attend to one of them (e.g., the green one, not the red one) and rate it for aesthetic quality. After a series of trials, they gave participants a surprise recognition test, found that recognition was at chance, and concluded that attention is needed to form new representations of shape. DeSchepper and Treisman's innovation was to add an implicit memory measure (response time) to the paradigm and to restructure the experimental procedure based on an inhibition mechanism model of attention. After a series of trials in which subjects attended to green shapes (ignoring red ones), target and distracter were reversed, i.e., a shape that previously appeared in red (distracter) now appeared in green (target). DeSchepper and Treisman report that previously ignored shapes were subsequently responded to more slowly than control shapes (never before exposed). This can only be accounted for by assuming that representations of these unattended novel shapes were formed in memory.

One could perhaps argue that the "unattended" shapes in these experiments were in fact attended and briefly noticed before being suppressed or that inhibition is by definition an attentional mechanism. However, it would be unwise to make such arguments, for this would raise the question of whether the noticing hypothesis could ever be falsified. Is seems, therefore, that there is now some good evidence for representation and storage in memory of unattended and not noticed novel stimuli.

There are limitations to these findings, however, that make it questionable how relevant such learning could be for SLA. The effect has been found so far only with visual perception and only when ignored stimuli compete directly and strongly for attention, which is generally not the case with second language input. There was no build up of memory strength with multiple presentation in the distracter role, making it unlikely that preconsciously established traces gradually build in strength until they are finally noticed. Not all subjects showed negative priming, suggesting strategy differences, but for those subjects who did demonstrate negative priming, a single act of attention was sufficient to change from inhibition to facilitation, which is necessary if knowledge is to
be available for use. This study did not assess subject awareness, but other studies have, and the general finding is that unaware subjects show negative priming of ignored stimuli, while aware subjects show facilitation (Driver & Baylis, 1993; Hasher, Stolzfus, Zacks, & Ryma, 1991; Neill & Valdes, 1992). Similar effects have been found in other studies contrasting conscious and unconscious perception (Merkle & Daneman, 1995). Unconsciously perceived stimuli can influence affective reactions; when the same stimuli are consciously perceived, these reactions are neutralized. Unconsciously perceived stimuli lead to automatic reactions; consciously perceived stimuli allow subjects to modify their reactions.

However interesting the finding of storage of nonconsciously stored novel information is theoretically, therefore, it appears to be of little benefit in language learning. In fact, if we are less able to access previously ignored than never-before-encountered information, we might have an explanation not for development in language learning but for non-learning, habituation of the self-instruction to ignore something.

**CONCLUSIONS**

Like most psychological constructs based initially on common experience, attention does not refer to a single mechanism but to a variety of mechanisms or subsystems, including alertness, orientation, detection within selective attention, facilitation, and inhibition. What these have in common (and do not share with the mechanisms of unattended, preconscious registration) is the function of controlling information processing and behavior when existing skills and routines are inadequate (Neumann, 1996). Learning in the sense of establishing new or modified knowledge, memory, skills and routines is therefore largely a side effect of attended processing. People learn about the things they attend to and do not learn much if anything about the things they do not attend to (Logan, Taylor, & Etherton, 1996).

The question of whether *all* learning requires attention remains problematic, and conceptual issues and methodological problems have combined to make a definitive answer illusive, even after a century of psychological experimentation. If the issue is seen as one of intention or the voluntary orientation of attention onto stimuli, the answer seems to be that intention is not a requirement. However, because goals and motivation are such important determinants of the focus on attention (Baars, 1986), paying deliberate attention to less salient or redundant aspects of L2 input may be a practical necessity. Since task demands are an equally important determinant of attentional focus,
instructional practices that focus learners’ attention on things that they are less likely to attend to or notice on their own also have a solid justification.

If the focus of inquiry is on what specifically in L2 input must be attended, there is a conflict between the encoding specificity hypothesis, which claims that only those stimulus attributes that are attended to in processing are encoded (Logan, Taylor, & Etherton, 1996), which implies that whatever evidence in L2 input is relevant to a particular learning domain must be attended, and the view that some aspects of L2 input are so subtle and abstract that they cannot possibly be attended to. The solution proposed in this paper is that attention must be directed to whatever evidence is relevant for a particular problem domain, i.e., that attention must be specifically focused and not just global, but that no abstract understanding of the significance of the evidence is required (although such understanding may be facilitative).

Finally, the important issue of whether there can be any learning (as opposed to activation of known information) on the basis of unattended, subliminal processing remains recalcitrant. Some recent evidence has been presented for it, but many psychologists have expressed the opinion that this dispute will never be settled, because zero-point questions are just not answerable (Baars, 1986; Merikle & Daneman, 1995). Baars suggests that the important question is not whether there can be any learning without attention and conscious involvement (unanswerable) but rather whether more attention results in more learning, to which the answer is clearly that it does. Kellogg and Dare (1989), who argue for both attended and unattended encoding, emphasize that while their conclusion that unattended encoding is possible, this “does not imply that unattended encoding has any practical value ... [since] the degree of elaboration resulting from unattended encoding appears to be too limited to have any substantive influence on human cognition or behavior” (p. 412).
REFERENCES


Richard Schmidt
Department of ESL
1890 East-West Road
Honolulu, Hawai‘i 96822

e-mail: schmidt@hawaii.edu