SECTION I

THEORETICAL ISSUES

1 *Attention*

Richard Schmidt

Introduction

The essential claim of this chapter is that the concept of attention is necessary in order to understand virtually every aspect of second language acquisition (SLA), including the development of interlanguages (ILs) over time, variation within IL at particular points in time, the development of L2 fluency, the role of individual differences such as motivation, aptitude and learning strategies in L2 learning, and the ways in which interaction, negotiation for meaning, and all forms of instruction contribute to language learning.

The theoretical issues dealt with in this chapter are organized around the basic assumptions, definitions, and metaphors concerning attention in psychology and include the idea of attention as a limited capacity, the notion of selective attention, and the role of attention in action control, access to awareness, and learning. It turns out that, like most psychological concepts initially based on common experience, attention is not a unitary phenomenon, but refers to a variety of mechanisms. 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, discussed towards the end of this chapter, indicates the possibility of some unattended learning, this appears limited in scope and relevance for SLA. There is no doubt that attended learning is far superior, and for all practical purposes, attention is necessary for all aspects of L2 learning.

A secondary goal of this chapter is to provide some of the details of the role of attention as that fits within a broader cognitive approach to understanding SLA, one that relies on the mental processes of language learners as the basic explanation of learning. 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.[1]

A full understanding of the ways in which awareness may shape SLA is beyond the scope of this chapter. 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 input results in more learning than can be reported verbally by learners. Knowledge of the grammar of one's L1 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 common phenomenon. In SLA, those most discussed at the present time are the Universal Grammar (UG) account, which argues for unconscious deductive reasoning from innate principles (Gregg, this volume), 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; MacWhinney, this volume). 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

[1] This chapter is a revised version of presentations at PacSLRF (Aoyama Gakuin University, Tokyo, March, 1998) and SLRF '97 (Michigan State University, East Lansing, October, 1997), under the titles 'The centrality of attention in SLA' and 'There is no learning without attention', respectively. The SLRF presentation was part of a point-counterpoint plenary with Jacquelyn Schachter of the University of Oregon, who presented 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. My own view is that conscious and unconscious processes probably interact in all domains of language, but that there is little evidence for learning without attention (one reading of 'unconscious') in any of them.

of learning, implicit and explicit, have to do with each other continues to be a topic of great debate within SLA and elsewhere. In SLA the question has frequently been posed in terms of whether or not 'learned' knowledge can become 'acquired' or whether the learner's conscious hypotheses can become internalized (Krashen, 1982; R. Ellis, 1993). 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), interact; to which the answer is probably yes, they do (Carr & Curran, 1994; N. Ellis, 1994a, 1996a, 1996b, this volume; Robinson, 1995b).

Since the concerns dealt with in this chapter concern the role of attention in SLA, it might be desirable to simply exclude all issues of awareness (Anderson, 1995). Unfortunately, it is probably impossible to separate attention and awareness completely, because of the common assumption that attention and awareness are two sides of the same coin (Carr & Curran, 1994; James, 1890; Posner, 1994), the emphasis in psychology on attention as the mechanism that controls access to awareness (discussed later in this chapter), and the reliance, in many experimental studies, on verbal reports as a method of assessing the allocation of attention. The solution adopted to this problem in this chapter is to limit the discussion of attention and its subjective correlate of 'noticing' to awareness at a very low level of abstraction. 'Noticing' is therefore used here in a restricted sense, as a technical term equivalent to 'apperception' (Gass, 1988), to Tomlin and Villa's (1994) 'detection within selective attention', and to Robinson's (1995b) 'detection plus rehearsal in short term memory.' My intention is to separate 'noticing' from 'metalinguistic awareness' as clearly as possible, 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 of which such instances may be exemplars. Although statements about learners 'noticing [i.e., becoming aware of] the structural regularities of a language' are perfectly fine in ordinary language, these imply comparisons across instances and metalinguistic reflection (thinking about what has been attended and noticed, forming hypotheses, and so forth), much more than is implied by the restricted sense of noticing used here.[2]

[2] As Truscott (1998) 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 appears necessary for understanding nearly every aspect of second and foreign language learning.

Understanding development

Some accounts of L2 development emphasize the importance of attention much 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 (which is not only unconscious but inaccessible in principle to consciousness), then the role of input is minimized and the role of attention to input even more so. Perhaps the only role for attention is that, presumably, at least the crucial evidence that triggers changes in the unconscious system must be attended (Schmidt, 1990). Connectionist models of learning, which are based on the processing of input and do not distinguish between competence and performance, also have little to say about attention, since input and output units are usually simply assumed to be attended. (Most connectionist 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 accounts of L2 development, especially those that are strongly psycholinguistic in approach (Bialystok, 1994; Carr & Curran, 1994; N. Ellis, 1994b, 1994c, 1996a; R. Ellis, 1997; Gass, 1988, 1997; Hatch, 1983a; Pienemann, 1989; Pienemann & Johnston, 1987; Robinson, 1995b; Skehan, 1998a; Swain, 1993, 1995; VanPatten, 1990, 1994, 1996; Wolfe-Quintero, 1992), within which attention to input is seen as essential for storage and a necessary precursor to hypothesis formation and testing. 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, frequency, the continuity of elements, and other factors that determine whether or not attention is drawn to them (Slobin, 1973, 1985; Towell & Hawkins, 1994). It has also been pointed out that attention is what allows speakers to become aware of a mismatch or gap between what they can produce and what they need to produce, as well as between what they produce and what proficient target language speakers produce (R. Ellis, 1994a; Gass, 1988, 1997; Schmidt & Frota, 1986; Swain, 1993, 1995, 1998).

Most discussions concerning the role of attention in L2 development focus exclusively on morphology and syntax, although a few

have dealt with lexical learning (N. Ellis, 1994b) and pragmatic development (Bialystok, 1993; Schmidt, 1993b). 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. However, if an utterance contains a pronoun, 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 as they map 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 gradually expanding processing space and the freeing of attentional capacity. For example, the crucial point for accurate production of third person singular – s is that the learner must have enough processing space available to generate a third person marker and keep it active in working memory until the appropriate moment arrives for attaching it to a verb.

VanPatten (1994) has argued that attention is both necessary and sufficient for learning L2 structure:

> 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 none the less 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)

Others who emphasize the importance of attention do not claim that attention is necessary for all learning. Carr and Curran (1994) claim that focused attention is required for some types of structural learning, but restrict this to cases where complicated or ambiguous structures are the object of learning. Gass (1997) argues against the principle that all L2 learning requires attention (attributing some learning to UG), but cautions that her arguments are not intended to weaken the claim that attention is important, merely to show that attention and awareness are not the only factors (p. 16).

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 accounts of the development of L2 fluency that are related to the psychological concept of automaticity (DeKeyser, this volume; 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 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 (N. Ellis, 1996a, this volume; Servan-Schreiber & Anderson, 1990). 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, an ability that develops as the result of age, experience, practice, and bilingualism (Bialystok, 1994a; Bialystok & Mitterer, 1987).

Understanding individual differences

Attention is a useful construct for understanding individual differences in SLA. As Tremblay and Gardner (1995) 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 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 – mediate between distant factors, including language attitudes and motivation, and achievement. Tremblay and Gardner found 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 else 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, short term or working memory capacity (Baddeley, 1986; Cowan, 1996; N. Ellis, 1996a, this volume; Harrington & Sawyer, 1992), is closely related to attention. Robinson (1995b) has suggested that my concept of 'noticing' can be redefined as detection plus rehearsal in short-term memory. Baddeley, Papagno, and Vallar (1988) have reported that such rehearsal is necessary for learning unfamiliar verbal material, although not necessary for forming associations between meaningful items that are already known. In the model of Baddeley (1986), there are three components of working memory: a 'central executive', explicitly related to attention and responsible for controlling the flow of information into working memory; a 'visuospatial sketchpad', a passive storage buffer for visual and spatial information; and an 'articulatory loop,' storing rich, detailed, and temporarily held information about the surface properties of language and allowing the relatively effortless recycling of the items currently in memory (Cowan, 1996). Another model relating attention to aptitude is that of Skehan (1998a), who suggests that the ability to notice what is in input is one of three factors in foreign language aptitude (see Sawyer & Ranta, this volume). The others are language analytic ability and the ability to retrieve chunks from memory to support fluent speech production.

Understanding the role of instruction

Sharwood Smith (1995) points out that input salience can be internally derived (when input becomes noticeable to the learner because of internal cognitive changes and processes) or externally derived (when input becomes more noticeable because the manner of exposure is changed). One major 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, 1994a; Hulstijn & de Graaff, 1994b; Long, 1988; Schmidt, 1990; Schmidt & Frota, 1986; Sharwood Smith, 1993, 1994; Terrell, 1991; Tomlin & Villa, 1994; 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, this volume; Doughty & Williams, 1998a; Skehan, 1996, 1998a).

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 non-native 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: 249).

If all these accounts are correct, attention is a crucial concept for SLA. The allocation of attention is the pivotal point at which learner-internal factors (including aptitude, motivation, current L2 knowledge, and processing ability) and learner-external factors (including the complexity and distributional characteristics of input, discoursal and interactional context, instructional treatment, and task characteristics) come together. What then happens within attentional space largely determines the course of language development, including the growth of knowledge (the 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: 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