A FORMAL INTEGRATED VIEW OF SPEECH, GESTURE, GAZE AND ITS IMPLICATIONS FOR LEARNING

Douglas W. Coleman
University of Toledo

LIEBERMAN ARGUES CONVINCINGLY (2000 *passim*) against modular linguistic theories which have created by assumption a distinct language organ. He cites Chomsky (1980a, 1980b, 1986), and Fodor (1983) as examples; in fact, the modularity assumption is still current, and has been restated or invoked repeatedly over the years since it was introduced by Chomsky (1964) over forty years ago. This has been made particularly clear in most theoretically-couched discussions of language learning, e.g. by Sharwood Smith (1985); White (1989); Cook (1993); Sharwood Smith (1994); Flynn, Martohardjono and O'Neil (1998); Chomsky (2000); Gass and Selinker (2001); and Carroll (2001). The neurological and behavioral evidence, however, shows the strict modularity assumption to be false. As Lieberman puts it, 'the neural bases of human language are intertwined with other aspects of cognition,[1] motor control, and emotion' (2000:2). The interdependence of these various networks has serious theoretical as well as practical implications.

In the present study I show how the interconnectedness of the various sensory-motor modalities is key to understanding how people actually learn to communicate via the neural mechanism of associative learning. In doing so, I show why certain input-based approaches in foreign language learning in fact provide adequate input while others do not.

1. A MODEL OF A 'NATURAL APPROACH' CLASSROOM INTERACTION. I start with an example of a simple communicative event—part of a Natural Approach (NA) foreign language lesson (see, e.g., Terrell 1977 *passim*) and show how one participant's linguistic behavior can be adequately modeled only by considering interconnected events across the modalities of speech, gaze, and gesture. The model is presented in the form of a Hard-Science Linguistics (HSL) **plex** (Yngve, 1996:171) of a **communicating individual** (Yngve 1996:124). The plex of a communicating individual is a representation of the physical-domain properties of a person relevant to his/her ability to communicate. Values of properties and the cause-effect relationships among them are formulated in terms of a network formalized using Boolean notation. Events are represented as Boolean pulses and conditions, as Boolean levels on connecting paths within the plex. I will show how, during a communicative event such as the one in question, these pertain *across* the sub-networks within the plex corresponding to the modalities of speech, gesture, and gaze.

A core component of an NA classroom linkage proceeds as in the following. A teacher (in this case, female) stands at the front of the room, next to a table, near the blackboard. The students sit facing her. On the table are several objects: a pencil, a book, and so on. She picks up a pencil, briefly making eye contact with various students, making sure she has their attention. Pointing at the pencil, she glances at it briefly, and articulates [tojestowuvek]. Laying the pencil back on the table, she picks up a book. She glances at the book while briefly pointing at it, but this time she articulates [tojestkśõška]. She returns the book to its place, and while simultaneously glancing down at the table and touching it with her hand, articulates [tojeststu:]. Next the teacher glances up, points at the ceiling, and articulates [tojestsufit].

2. ASSOCIATIVE LEARNING IN THIS SITUATION. In the segment of Natural Approach classroom linkage just described, there are several iterations of the same linkage tasks.

Each time, a correlated set of events has occurred involving speech, gesture, and directed gaze. The teacher has brought an object into salience for the students by means of gesture and directed gaze. She has performed a similar articulation—[tojest]; from this, there is **at the neurological level internal within the students**[2] an association formed between the gesture(s), directed gaze, and this speech event.

In addition, some events have occurred which distinguished each iteration. The teacher articulated [owuvek] while glancing at / holding up / pointing to the pencil, [kśõška] while glancing at / holding up / pointing to the book, [stu:] while glancing at / touching the table, and [sufit] while glancing at / pointing to the ceiling. From this, an internal association will occur between the neurological events resulting from the perception of each articulation [owuvek], [kśõška], [stu:], or [sufit] with the perception of anything in the same cognitive-perceptual category as the pencil, book, table, or ceiling, respectively.[3]

3. A (PARTIAL) FORMAL DESCRIPTION OF THE EVENTS. The following description, given the inherent limitations of space, is necessarily presented only in overview.

Figure 1 shows—in HSL Boolean network notation—the portion of the teacher's procedural properties relevant to her ability to take part in the portion of the NA classroom linkage of particular interest here. A pulse enters on the line at left and initiates the task procedure <Establish Attention>,[4] goes from there to the task procedure <Establish Salience>, from there to the task procedure <Name>, and from there to the task procedure <Release>. The task procedure <Establish Attention> has subtasks in the perceptual system by which the teacher establishes that students are attending to her. The task procedure <Establish Salience> sends a pulse (σ-pul) that initiates subtasks in motor-sensory systems related to speech and gesture. The task-active line of <Establish Salience> is a conditional property (Prop-Directed-Gaze-lev) relevant to motor-sensory processes controlling directed gaze; when <Establish Salience> is active, this property is TRUE.[5] Feedback from the gesture and speech subsystems, respectively, is in the form of subtasks-done pulses ϕ1-pul and ϕ2-pul (connected by an AND-collector: an AND gate which does not need to receive its pulses simultaneously). The subtasks-initiating pulse ν-pul and the subtasks-done pulse ϕ3-pul of <Name> are to/from the speech subsystem, while those of <Release>—ρ-pul and ϕ4-pul—are to/from the gestural subsystem. The output pulse branch ω-pul provides a signal to the directed-gaze subsystem, allowing its release from linkage-related tasks (see the description accompanying **Figure 2**). The diagram in **Figure 1** describes the overall events in the

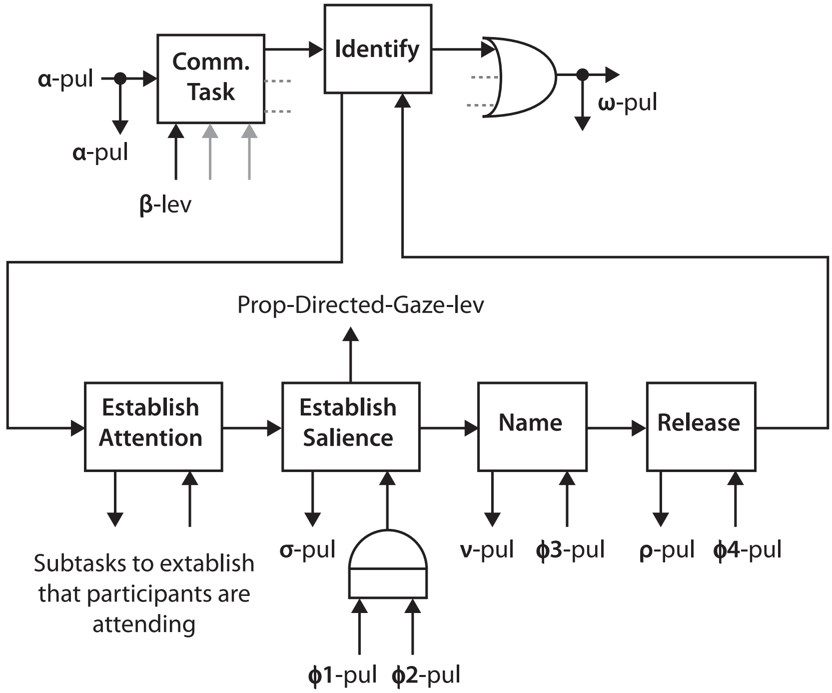

Figure 1. *Higher-order procedural properties in teacher plex.*

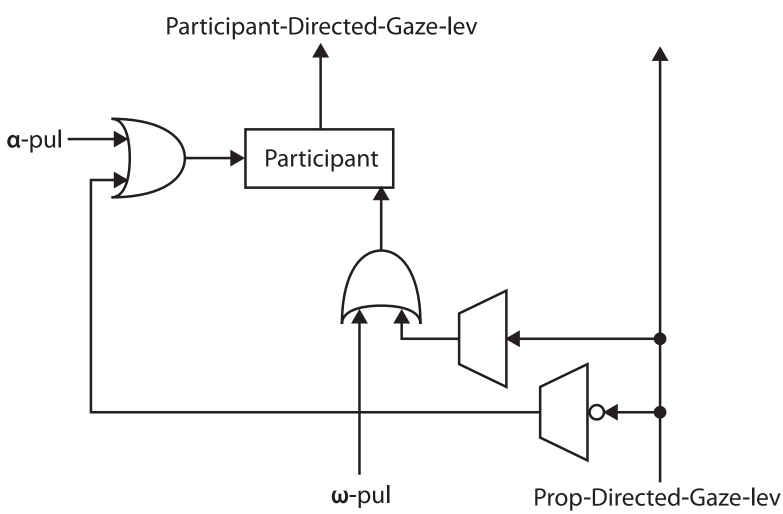

Figure 2. *Part of the directed-gaze subsystem of the teacher's plex.*

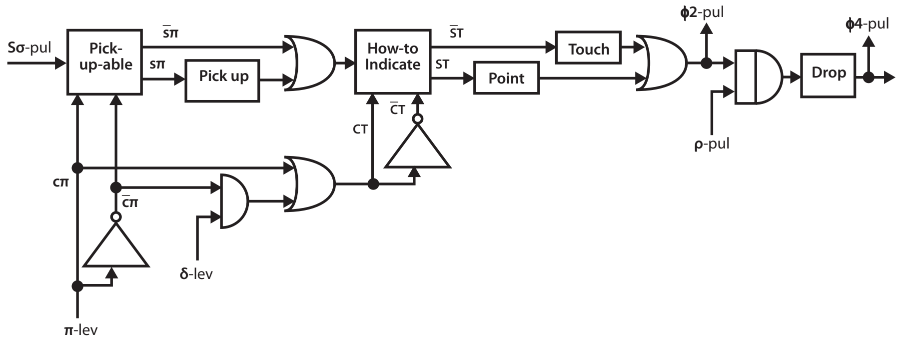

Figure 3. *Part of the gestural subsystem of the teacher's plex.*

teacher's plex relevant to her speech, gesture, and directed gaze during this short linkage segment, as they are described in the earlier narrative.

Figure 2 shows relevant procedural properties from the directed-gaze subsystem of the teacher's plex. The input pulse which initiates <Establish Attention> in **Figure 1** is synchronized with α-pul, shown here in **Figure 2**. The pulse passes on to <Participant>, which functions as a pulse-to-level converter. When Prop-Directed-Gaze-lev undergoes a transition from Boolean FALSE to TRUE, a level-to-pulse converter (the upper-left of two gates shaped like trapezoids in **Figure 2**) passes a pulse input through an OR gate to the input on the bottom right of <Participant>, temporarily resetting Participant-Directed-Gaze-lev to FALSE. When Prop-Directed-Gaze-lev, on the other hand, undergoes a transition from Boolean *TRUE* to *FALSE*, a second level-to-pulse converter (the lower-right of two gates shaped like trapezoids in **Figure 2**, the one with a circle indicating its input is inverted) passes a pulse input through an OR gate to the input on the left side of <Participant>, resetting Participant-Directed-Gaze-lev back to TRUE. Participant-Directed-Gaze-lev remains at Boolean TRUE until there is a pulse ending the linkage segment normally (ω-pul).

Figure 3 shows relevant procedural properties from the gestural subsystem of the teacher's plex. Sσ-pul is connected (with some additional elements in between) to σ-pul in **Figure 1**. Most of the logic of **Figure 3** is devoted to showing the interactions between the conditions π-lev (TRUE if the teacher's cognitive system judges the prop being brought into salience able to be picked up) and δ-lev (TRUE if the prop is too distant to be picked up). Of particular relevance here, however, are ρ-pul and ϕ2-pul, at the right of the diagram, which also connect the procedural properties of gestural subsystem to the higher-order procedural properties shown in **Figure 1**. The connection via ϕ4-pul is discussed later.

Figure 4 shows relevant procedural properties from the speech subsystem of the teacher's plex. SAσ-pul is connected (with some additional elements in between) to σ-pul in **Figure 1**. The task <To> has subtasks (not shown) which handle the portion of the articulation represented above as [to]. When this task is completed, a subtasks-done pulse (ϕ1-pul) is returned to the higher-order task properties in **Figure 1**. As soon as ϕ1-pul (from the speech system—**Figure 4**) and ϕ2-pul (from the gestural system—**Figure 3**) have both been received by the higher-order task procedure <Establish Salience> (**Figure 1**), the

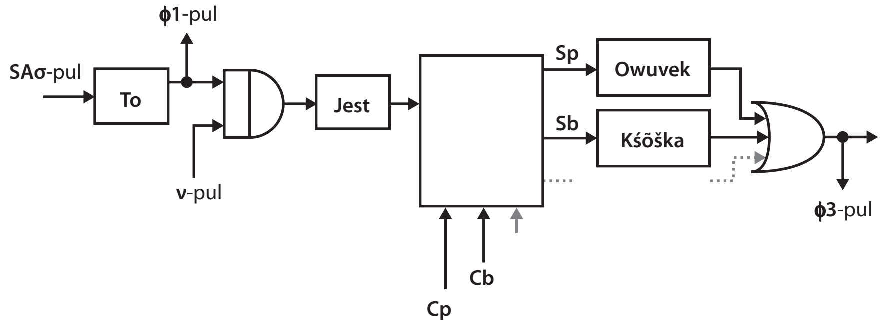

Figure 4. *Part of the speech subsystem of the teacher's plex.*

higher-order task <Name> initiates, sending an initiate-subtasks pulse on ν-pul (**Figures 1** and **4**). At this point, a pulse will proceed in the speech subsystem to the task <Jest>, whose subtasks will result in the articulation represented above as [jest]. When <Jest> is done, a pulse continues on to the selection procedure in the center of **Figure 4**. For the sake of simplicity, only two conditional properties are shown, Cp and Cb, where the p and b stand for speaker reference to pencil and book, respectively. Depending on which conditional property is Boolean TRUE, the corresponding output line (Sp and Sb) will carry a TRUE pulse. The immediate outcome will be that either the task <Owuvek> or the task <Kśõška> will be initiated, in turn resulting in either the articulatory events represented as [owuvek] or [kśõška], respectively. Regardless of which articulatory task is performed, a pulse passes on to the OR gate at the far right of **Figure 4**, causing a subtasks-done feedback pulse to be sent back to the higher-order task <Name> via ϕ3-pul (**Figures 1** and **4**).

Note that it is the <Drop> task in the gestural subsystem (**Figure 3**) which returns a subtasks-done pulse via ϕ4-pul to the higher-order task <Release> (**Figure 1**), effectively concluding an iteration of the linkage portion.

What this model is intended to demonstrate is that the intimate and—highly relevant—interdependencies among the speech, gestural, and directed gaze subsystems can be formally described, using HSL Boolean network notation.

4. THE CRITICAL VALUE OF THE CORRELATED EVENTS. Twenty years ago, Klein (1986:44) made the following observation:

> Suppose you were locked in a room and were continually exposed to the sound of Chinese coming from a loudspeaker; however long the experiment continued, you would not end up speaking Chinese...

As Klein's Chinese Room anecdote illustrates, the 'input' for the students cannot consist of Chomsky's (1964:26) 'primary linguistic data'.[6] The input for learning to communicate—what we conventionally call language learning—includes 'the information received *in parallel* to the linguistic input in the narrower sense (the sound waves [of speech])' (Klein 1986:44). Thus, it is clear that we should be integrating into our formal descriptions such information as is included above, especially if we are to understand how people actually learn to communicate.

5. TWO 'INPUT-BASED' APPROACHES. Terrell's (1977) Natural Approach incorporates what is needed to prevent the classroom from being Klein's (1986) Chinese Room. In the Natural Approach, the perception of speech or writing does not occur in isolation from people communicating, but is linked to gestural and directed gaze behaviors of the teacher (and sometimes of the students) and to objects present in their environment.

Now consider another supposedly input-based approach: the 'grammar-processing approach' of VanPatten (1996:2). In discussing the role of input, VanPatten clearly accepts Chomsky's assumption of input as consisting of the primary linguistic data. VanPatten (1996:6)states, 'The definition of input is limited to meaning-bearing input, language that the learner hears or sees that is used to communicate a message.' Note that meaning in VanPatten's view is not something that depends on the properties of people (which we might represent in terms of the plex of a communicating individual) or their interactions (which we might represent as linkage properties), but is something that can exist apart from people, borne by input—which purportedly consists of language. He elaborates quite a bit on this point, identifying input as 'restricted to samples of second language that learners hear or see to which they attend for its propositional content' (VanPatten 1996:10). He assumes that the input consists of words and grammar. But if it did, then we would learn Chinese in Klein's Chinese Room.

There are some immediately significant consequences for VanPatten's input-processing theory. For example, he draws up a number of principles for his theory, such as 'P1(a): Learners process content words in the input before anything else' (VanPatten 1996:18).[7] The reader is cautioned that if this principle appears self-evident, he (or she) should keep in mind that—in any real-world sense—the input actually consists of sound waves, light waves, kinetic energy, and the like, all of which impinge on the learner's body and cause internal changes of state describable in physiological (and especially neurobiological) terms. Words, on the other hand, are mental objects, which therefore exist only in the logical domain (and thus in a person's conscious subjective experience). It is curious that VanPatten never notices that in order for P1(a) to make sense, the learner somehow needs to know what the words in the input are (supposedly in an unknown language); nor does he take up the issue of how the learner could know which words in the input are content words and therefore should be processed first; nor does he explain how input purportedly consisting of samples of language unknown to the learner can be meaningful to him/her; and so on. In short, from a theoretical perspective, input processing suffers from the presence both of unsupported assumption and self-contradiction.

In his chapter 'Processing Instruction', VanPatten (1996:55–86) presents a number of examples of activities for processing instruction which

> utilize what is best termed 'structured input.' 'Input' refers to the fact that during the activities, learners do not produce the targeted grammatical form or structure. Instead, they are engaged in actively processing input sentences. (63)

Actividad J: Los parientes. What are the things that relatives do to us? They can bother us, visit us, criticize us, love us, and so on.

Paso 1. Read each statement and select the ones that you think are typical.

Los parientes...

a. ☐ *nos molestan.*
b. ☐ *nos critican.*
c. ☐ *nos ayuden* (help).
d. ☐ *nos visitan.*
e. ☐ *nos quieren* (***querer*** = to be fond of).
f. ☐ nos ___________.

Figure 5. *A 'structured input activity' (VanPatten 1996:66).*

Figure 5 shows a typical example from VanPatten's book (ibid:66). This activity is among those explicitly described as providing 'structured input' (ibid:63–64). In it, we see three different types of translation cues. The first is contained within what purport to be the instructions for the activity ('They can bother us, visit us, criticize us, love us, and so on'); this is directed specifically at *molestan*, *critican*, and *visitan* in (a), (b), and (d), respectively, three items of the activity not cued within the activity itself. Clearly, this is to prevent the false cognate syndrome in the first case and to tell students that the other two are true cognates. The other two complete items—(c) and (e)—are translation-cued within the activity itself, albeit in slightly different ways; this is almost certainly attributable to the grammatical irregularity of *quieren* in item (e). The presence of these cues implicitly admits the falsehood of the assumption that input consists of second language primary linguistic data in Chomsky's terms.

7. CLOSING REMARKS. VanPatten claims the input for second language learning consists of sentences containing words arranged in grammatical structures carrying meaning. If this were the case, the Chinese Room approach would work.

But I repeat, as Klein (1986) pointed out, it does not.

The assumption that input consists of the primary linguistic data is false. Rather, the input is the totality of a learner's sensory experience. Therefore, the so-called Poverty of Stimulus does not exist (see Coleman 2005).

An approach like that of VanPatten's must tacitly acknowledge its entry into the Chinese Room (typically via the presence of translation cues and similar hints), even if it makes explicit claims to the contrary.

In an input-based approach that works, one which provides adequate input for learning how to communicate, speech or writing must be viewed in terms of communicative interactions (linkages) that include correlated events between/among 'the linguistic input in the narrower sense (the sound waves [of speech])' and 'the information received *in parallel*' to it (Klein 1986:44).

This is because traditional, language-focused approaches to understanding how people communicate formalize only the mental objects of grammar (grammar in the broad sense), which exist only in the logical domain. But as shown above, there are real-world (physical-domain) events and properties of people relevant to our understanding of the situation. Because people and the communicative events they are involved in exist in the physical domain, traditional language-based approaches (which confine themselves to the logical domain) permit only vague and imprecise appeals to extra-linguistic context. Let us suppose that what the field of linguistics has as its ultimate goal is a useful, scientifically-based model of how people communicate. If so, we must turn away from the logical domain and accept that it is not language, but what we marginalize as context (people communicating in their surroundings) that properly concerns a science of linguistics.

---

1 It is clear from the discussion that follows in Lieberman (2000:3–11) that he includes under 'other aspects of cognition' neural processes associated with sensory input.

I refer just above to the **strict** modularity assumption because, as one anonymous reviewer of this paper helpfully pointed out, 'the neurological evidence also shows that neurons performing tasks with similar functions are grouped in a kind of natural efficiency'. Indeed, this is one of Lieberman's key points—and that these groupings in no way constitute parts of a 'language organ'. For example, '[a] particular structure, such as the putamen (part of the subcortical basal ganglia), can control similar sequencing operations in manual movements [and] speech production', i.e. both non-linguistic and linguistic functions (Lieberman 2003:10).

2 The following is to insure that there is no misunderstanding: this internal association at the neurological level is not to be equated with an external S→R association as conceived of by, say, Bloomfield (1933). Behaviorist S→R theory, in fact, denies the relevance of not-directly-observable internal processes, confusing the not-directly-observable with the abstract. For discussion, see Coleman (2004).

3 Obviously, the cognitive category with which the perception of any pencil, book, table, or ceiling is attached depends on each student's internal properties, a result in part of individual prior experience. If the student lacks the neurological structure associated with a cognitive-perceptual category 'pencil', then perception of [owuvek] will be associated with whatever cognitive-perceptual category pre-exists for the perception of the pencil in the teacher's hand. I am thinking, of course, of cases in which the student's prior experience may have resulted in a cognitive-perceptual category which includes pencils but not mechanical pencils; one which includes pencils, chalk, charcoal, but not pens or brushes; etc.

4 When I refer to a line being at a certain level, I mean it is stable at Boolean TRUE or FALSE in contrast to a line carrying a pulse, which is typically FALSE until it only momentarily goes to TRUE, then reverts to FALSE. When I say that a task procedure is 'initiated' by an input pulse, I mean that (a) an output pulse is sent to its subtasks and (b) its task-active state becomes Boolean TRUE.

5 For the reader's convenience, individual line labels end with either 'pul' or 'lev' as a reminder as to whether they carry a Boolean TRUE pulse or TRUE / FALSE level.

6 Coleman (2005) discusses in detail the relevance of the Chinese Room anecdote and demonstrates, based on a number of reasons, why the input cannot be thought of as 'the primary linguistic data'.

[7] In linguistics (in this case an applied subfield), we often see such an enterprise being carried out riddled with principles which are posited and then taken as given. In such cases, it should immediately be obvious that the enterprise is in fact being conducted as philosophy, rather than science. In the latter domain, we instead see hypotheses offered up for testing in the real world.

REFERENCES

Bloomfield, Leonard. 1933. *Language*. New York: Holt.

Carroll, Susanne E. 2001. *Input and evidence: The raw material of second language acquisition*. Philadelphia: John Benjamins.

Chomsky, Noam. 1964. *Current issues in linguistic theory*. The Hague: Mouton.

———. 1980a. Initial states and steady states. In *Origins and evolution of language and speech*, ed. by H. B. Steklis, S. R. Harnad & J. Lancaster, 46–57. New York: New York Academy of Sciences.

———. 1980b. Rules and representations. *Behavioral and brain sciences* 3:1–61.

———. 1986. *Knowledge of language: Its nature, origin, and use*. New York: Praeger.

———. 2000. *New horizons in the study of language and mind*. New York: Cambridge University Press.

Coleman, Douglas W. 2004. Linguistic meaning in the physical domain. *LACUS forum* 30:331–41.

———. 2005. Language learning input and input for learning to communicate. *LACUS forum* 31:203–13.

Cook, Vivian. 1993. *Linguistics and second language acquisition*. New York: St. Martin's.

Flynn, Suzanne, Gita Martohardjono & Wayne O'Neil (eds.) 1998. *The Generative study of second language acquisition*. Mahwah NJ: Lawrence Erlbaum.

Fodor, Jerry. 1983. *Modularity of mind: An essay on faculty psychology*. Cambridge MA: MIT Press.

Gass, Susan M. & Larry Selinker. 2001. *Second language acquisition: An introductory course*. Mahwah NJ: Lawrence Erlbaum.

Klein, Wolfgang. 1986. *Second language acquisition*. New York: Cambridge University Press.

Lieberman, Philip. 2000. *Human language and our reptilian brain: The subcortical bases of speech, syntax, and thought*. Cambridge MA: Harvard University Press.

———. 2003. Language evolution and innateness. In *Mind, brain, and language: Multidisciplinary perspectives*, ed. by Marie T. Banich & Molly Mack, 3–22. Mahwah NJ: Lawrence Erlbaum.

Sharwood Smith, Michael. 1985. From input to intake: On argumentation in second language acquisition. In *Input in second language acquisition*, ed. by Susan M. Gass & Carolyn Madden, 394–403. Rowley MA: Newbury.

———. 1994. *Second language learning: Theoretical foundations*. New York: Longman.

Terrell, Tracy D. 1977. A natural approach to second language acquisition and learning. *Modern language journal* 61:325–36.

VanPatten, Bill. 1996. *Input processing and grammar instruction in second language acquisition*. Norwood NJ: Ablex.

White, Lydia. 1989. *Universal grammar and second language acquisition*. Philadelphia: John Benjamins.

Yngve, Victor H. 1996. *From grammar to science: New foundations for general linguistics.* Amsterdam: John Benjamins.