Introduction

First exposure, input processing, and theorizing

ZhaoHong Han and Rebekah Rast

Since Corder's (1967) seminal postulation that intake does not equal input, the construct of intake has taken on central importance in second language acquisition (SLA) research, serving increasingly as a linchpin tying together a heterogeneous spectrum of theoretical and empirical endeavors. Among the core concerns to the researchers are the following: (a) What induces as well as constrains intake, both externally and internally? (b) What form may intake assume? (c) Can intake be externally manipulated (e.g. through pedagogical intervention)? Although the general understanding of the input–intake asymmetry has come a long way, it also seems hampered – especially where (c) is concerned – by an, as yet, very limited body of knowledge vis-à-vis learners' own intake capability, including their 'default' approaches, natural inclinations, and available strategies.

Input processing (i.e., the mediating process of intake) on first exposure to a second language has in recent years garnered increasing research attention, as is evident in the two special issues that recently appeared in the scholarly journals of *Language Learning* (Gullberg & Indefrey 2010) and *Second Language Research* (Carroll 2013). Previously, the literature has only sporadically seen publications on the topic (see, e.g., Han & Peverly 2007; Rast 2008).

Driving the current research interest has been, among other things, the desire to understand the natural processes and mechanisms underlying a learner's capacity to break into the 'wild,' including how much input learners can process, and how they go about processing that input. Such understanding is critical to both theorizing and theory-guided empirical SLA research, as, to date, much of our understanding on input processing has derived from assumptions and facts about learners *en route* or *al fine*. Not including learners at the initial stages would necessarily result in theoretical and empirical inadequacy, but, more importantly, an incomplete picture of second language acquisition.

The present volume further contributes to the growing line of research on first exposure, providing a collection of studies that examine various aspects of processing, using a plethora of methodologies. While the entire domain of

research on first exposure has yet to see the emergence of a comprehensive theoretical framework, the studies included in this collection draw from a variety of theoretical perspectives, but mainly three: VanPatten's (1996) input processing theory, Carroll's (2001) autonomous induction theory, and usage-based theory (see, e.g., Ellis 2008).

Guided by VanPatten's (1996) input processing theory, Chapter 1 (Han & Sun) and Chapter 2 (Park) both sought to verify among first exposure learners the Primacy of Meaning Principle. This principle stipulates that second language learners selectively pay attention to meaning and form during input processing and that they prioritize meaning over form. A specific corollary, VanPatten suggests, is that learners process content words before processing grammatical morphemes. The studies reported in Chapter 1 and Chapter 2, though carried out in different contexts involving different target languages (Norwegian and Korean), both provide counter-evidence to this claim. Both studies found that first exposure learners start out with a form-oriented rather than a meaning-oriented approach to input processing, concluding that in order to assume a meaning-oriented approach, as articulated in VanPatten (1996, 2004, 2007), learners would have developed some proficiency in the target language as a prerequisite. In other words, a meaning-based approach, the authors contend, presupposes some experience with and knowledge of the target language. Carroll (2013) usefully differentiates *meaning* into *reference*, which can be derived from a word in isolation (as when the referent is clear in the immediate environment), and *sense*, which is derived in concert with other linguistic elements (as in a phrase or a sentence). In this light, meaning-based processing of words in continued discourse, such as a listening or reading text, would be possible only when the learner has sufficient knowledge of the words and their 'neighbors.' In the case of first exposure learners, such knowledge, with the possible exception of cognates (cf. Carroll 2012; Rast 2010), is not yet available (see also Han & Liu 2013), a critical difference between *ab initio* learners and learners *en route* or *al fine*.

Within a usage-based framework, Chapter 3, by Rast and colleagues, investigates the learning of inflectional morphology in first exposure learners of Polish. The study is unique in that it focuses on documenting the input given to the learners. The learners' ability to perceive elements in the input and use these elements is then inferred from measurements of learners' recognition and production via a grammaticality judgment and a sentence production task. The data were analyzed by cross-checking input properties (frequency and lexical transparency) and learner performance. Levels of processing were ascertained by repeating the same tasks at various intervals. The authors evaluate, in particular, the facilitating effects of frequency and transparency, and their findings confirm that familiarity with target stems helps learners process novel inflectional forms in the input. This chapter also

looks into the relationship between the learning outcomes and individual differences in learning and cognitive styles.

In keeping with her reasoning along the lines of autonomous induction (Carroll 2001), Carroll, in Chapter 4, examines word processing in first exposure learners and beginners of German. She claims that in first exposure learners, word processing does not happen in quite the same way as it does in beginners, due to the fact that a first exposure learner does not have morphosyntactic knowledge of the target language but a beginner does, though to a limited extent. Word processing in the absence of any morphosyntactic knowledge is necessarily confined to encoding sound-referent associations of words, but having morphosyntactic knowledge would enable word processing at sentence or even discourse level, thereby deriving individual sense from each sentential element to contribute to the total meaning of a larger linguistic unit. The study subjects its participants to a particular word training paradigm involving cognates and compounds, the results showing, *inter alia*, that cognates are easy to segment and learn, which extrapolates to the understanding that for familiar items, frequency or repeated exposure is *not* necessary for successful word learning.

Assuming that first exposure learners are different from beginners, with the former having zero knowledge of the target language and the latter possessing incipient knowledge, Sagarra, in Chapter 5, claims that an understanding of how beginners process input is just as important as first exposure studies in the identification of developmental continuity or lack thereof. Sagarra focuses on an emerging finding from first exposure research, namely that first language (L1) transfer does not seem to occur in initial input processing (Han & Liu 2013; Park & Han 2008), investigating the issue of morphological transfer in beginners. The study controls for cross-linguistic difference in inflectional morphology using data collected from native-speaker control groups – monolingual speakers of languages that are morphologically rich (Romanian and Spanish) or poor (English) – and English and Romanian beginning learners of Spanish. Eye-tracking data from a reading and picture verification task measured early processing with first-pass duration on S(subject) and V(verb) in sentences exhibiting SV agreement or disagreement as well as comprehension. Results found no evidence of L1 transfer in beginners. In the author's words, "beginning learners are immune to morphological transfer effects." This study, therefore, argues in favor of little to no L1 transfer of inflectional morphology during the developmental phase from first exposure to beginning L2 learners.

These five empirical studies (reported in Chapters 1–5) exhibit epistemological and methodological differences. The studies by Han and Sun and by Park seek to understand learner spontaneous processing of input and accordingly employ uncontrolled input as the stimulus and subsequently probe

learners' 'reactions.' This line of inquiry will ultimately shed light on what learners can process on their own. This type of understanding is sorely needed to serve as a basis for pedagogical intervention, for, as Corder (1967) has reminded us, we cannot teach language; we can only create conditions in which it will develop spontaneously in the learner's mind in its own way.

In the studies reported in Chapter 3 (Rast et al.) and Chapter 4 (Carroll), the primary interest is in assessing the impact of input properties on processing. The input is therefore controlled, and processing is arguably aided by making word meaning available to the learner, either through instruction (Rast et al.) or training trials (Carroll). Studies of this nature (along with laboratory studies on input processing of artificial or semi-artificial languages) promise to lead to an understanding of what information in the input (e.g., frequency, cognate status) is relevant to processing, how it is utilized during processing, and what outcomes of processing it enables.

Similarly, in the Sagarra study (Chapter 5), input is highly manipulated such that different processing conditions are created, the goal being to track down the function of a given factor, namely L1 transfer, in input processing. Studies of this nature are critical to testing existing hypotheses and necessary to honing an understanding of individual factors that have been identified in previous studies.

Input processing research after all deals with an invisible phenomenon. Probing the unseen has been notoriously difficult, and how to measure processing is a particularly vexing issue. Chapter 6, by Moreno, delves into these concerns, urging that input processing studies must pay attention to their internal validity. Moreno takes a close look at methodology in VanPatten's (1990) study of allocation of attention to form and meaning, as well as in several replications of his study. To ensure the internal validity, Moreno offers recommendations, including learner think-alouds and a tighter control over variables such as frequency and modality (i.e., written versus aural).

Following the two aforementioned special issues on the topic in *Language Learning* and *Second Language Research*, this collection of studies offers yet another lens through which to view results based on empirical studies of first exposure, methodologies used to obtain these results, the engagement such results might find with existing theories of input processing, and directions for the future. To further provoke and unsettle this domain of research, we invited Bill VanPatten to write an epilogue for the present volume. As so desired, VanPatten critically discusses the findings from the studies, taking the unique opportunity to clarify his own thoughts on input processing (e.g., his current conception of 'form'), pointing out conceptual and methodological issues that have yet to be resolved in input processing research, and even stirring further controversies in input processing research (or SLA research in general, for that matter). One such point of potential controversy concerns the definition

of 'processing,' which VanPatten feels adamantly should entail connecting a form with a meaning – even right from the start. This essentially rejects the idea that there can be levels, or phases, of processing (Carroll 1999; Chaudron 1985). Another point he makes and one worth mentioning here is that processing should be dissociated from noticing given that awareness is not a requirement for processing. Implied in VanPatten's position is that input processing may be entirely implicit and unconscious. For us, this immediately triggers many questions, among them the following: (a) If noticing (Schmidt 1990) is not implicated in processing, does that mean that learners do not consciously employ strategies during input processing? If they do, however, how would we be able to observe these strategies other than by eliciting learners' own protocols? (b) If input processing does not entail noticing, then how do we account for all the 'processes' required of learners in order to create representations that allow for form–meaning mapping, such as perceiving sounds in the speech stream and segmenting speech in order to identify 'words' that can be comprehended? (c) If input processing is only implicit and unconscious, then how do we account for all the explicit and conscious work learners do on their input, such as hearing a word and repeating it (regardless of whether they understand what they are repeating)? Are they not 'processing' in this act of listening and pronouncing a group of novel sounds? (d) Is input processing entirely unconscious? How do we know? (e) If input processing is partially conscious and partially unconscious, what methodologies would illuminate the hybridity? Using methodologies that do not probe the learner's perspective on processing does not prove the unconscious nature of input processing; by the same token, tapping into learner protocols should not lead to the conclusion that processing is entirely conscious.

Clearly, much needs to be further researched, and it is our hope that this volume offers ideas and insights that will lead to a substantiation of current understandings. Still, as it stands, the present book offers a lens through which readers can gain not only an understanding of input processing in first exposure learners but also many useful insights into other more general concerns, such as how to investigate frequency effects, a key construct in current usage-based approaches in applied linguistics research, not to mention ideas on many other possible research topics.

REFERENCES

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

(2001). *Input and Evidence: The Raw Material of Second Language Acquisition*. Amsterdam: John Benjamins.

(2012). Segmentation on first exposure to an L2: Evidence for knowledge-driven, top-down processing. In K. Braunmüller, C. Gabriel, and B. Hänel-Faulhaber

(eds.), *HSM13 Multilingual Individuals and Multilingual Societies* (pp. 23–45). Amsterdam: John Benjamins.
(2013). Introduction to the special issue: Aspects of word learning on first exposure to a second language. *Second Language Research*, *29*(2), 131–144.
Chaudron, C. (1985). Intake: On models and methods for discovering learners' processing of input. *Studies in Second Language Acquisition*, *7*, 1–14.
Corder, S. P. (1967). The significance of learners' errors. *International Review of Applied Linguistics*, *5*, 161–170.
Ellis, N. C. (2008). Usage-based and form-focused language acquisition: The associative learning of constructions, learned attention, and the limited L2 end state. In P. Robinson and N. Ellis (eds.), *Handbook of Cognitive Linguistics and Second Language Acquisition* (pp. 372–405). New York: Routledge.
Gullberg, M. and Indefrey, P. (eds.) (2010). The earliest stages of language learning. *Language Learning*, *60*, Supplement 2, 1–283.
Han, Z.-H. and Liu, Z. (2013). Input processing of Chinese *ab initio* learners. *Second Language Research*, *29*(2), 145–164.
Han, Z.-H. and Peverly, S. (2007). Input processing: A study of *ab initio* learners with multilingual backgrounds. *The International Journal of Multilingualism*, *4*(1), 17–37.
Park, E. S. and Han, Z.-H. (2008). Learner spontaneous attention in L2 input processing: An exploratory study. In Z.-H. Han (ed.), *Understanding Second Language Process* (pp. 106–132). Clevedon: Multilingual Matters.
Rast, R. (2008). *Foreign Language Input: Initial Processing*. Clevedon: Multilingual Matters.
(2010). The role of linguistic input in the first hours of adult language learning. *Language Learning*, *60*, Supplement 2, 64–84.
Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, *11*(2), 129–158.
VanPatten, B. (1990). Attending to form and content in the input. *Studies in Second Language Acquisition*, *12*, 287–310.
(1996). *Input Processing and Grammar Instruction: Theory and Research*. Norwood, NJ: Ablex.
(2004). *Processing Instruction*. Mahwah, NJ: Lawrence Erlbaum Associates.
(2007). Input processing in adult second language acquisition. In B. VanPatten and J. Williams (eds.), *Theories in Second Language Acquisition* (pp. 115–136). Mahwah, NJ: Lawrence Erlbaum Associates.

1 First exposure: a replication of Han and Peverly (2007)

ZhaoHong Han and Yayun Anny Sun

Over the four decades of second language acquisition (SLA) research, much has been said and done about input (Ellis 2002; Gass 1997; Gass & Madden 1985; Long 1996), but surprisingly little on input processing. As Carroll has observed, "most research dealing with 'input' provided descriptions of what people say to learners, not what learners can perceive and represent" (2005: 81), and "[input processing] remains one of the most under-theorized and under-researched areas of our field" (1999: 338). This is unfortunate, given the nature of second language acquisition as fundamentally cognitive (Doughty & Long 2003; Long 2009). However, the scenario is changing, slowly but surely. In recent SLA literature, we see not only mounting research ostensibly relating to processing, both input- and output-based, but also a shift of attention from representation to processing, most acutely within the generative paradigm.

Where research on input processing is concerned, the focus of this chapter, mainstream efforts have by and large been funneled into sentence-level processing (e.g., Bates & MacWhinney 1981; MacWhinney 2001; Marinis et al. 2005; Roberts 2007);[1] in contrast, attempts to understand the processing of discourse-level input, not to mention processing of such input at the initial state of SLA, are as yet sparse and disconnected (cf. Bremer et al. 1996; Singleton & Little 1984). It is the purpose of this chapter to help narrow the gap, and, in concert with all other chapters of this volume, to do so by focusing on the input processing in learners who are exposed to the target language input for the very first time.

In what follows, we report on a replication of Han and Peverly (2007), which sought to examine the contents and manner of input processing in so-called *ab initio* learners. We begin by briefly reviewing the theoretical background of the study. We then present the methodology and results,

We would like to thank the reviewers for their very perceptive critique and helpful suggestions on an earlier draft of this chapter. All remaining faults are ours.

[1] Research on sentence processing has mostly focused on uncovering strategies, such as garden-pathing, employed in resolving locally ambiguous input and on parsing failures.

followed by a discussion of the findings, especially in relation to the original study. We conclude with a brief discussion of the significance and limitations of the study and future research directions.

Theoretical background

Input–intake in SLA

Input, by all theoretical accounts, is a staple condition of second language acquisition; if there is no input, there is no acquisition. As such, input has received abundant attention in SLA research. Even a cursory look at the literature shows that definitions of input alone have run the gamut from generic to technical ones. A generic definition of input can be found in Chaudron (1985: 3): "The input available to second language learners is the raw data from which they derive both meaning and awareness of the rules and structures of the target language."

A technical definition of input is given in Carroll (1999: 337), where "input equals objective properties of the stimulus array less the effects of selective attention (the intake)."

Input serves to exemplify what the target language is like and how it is used in the service of communicative functions and purposes. It potentially provides the data a learner needs to formulate, confirm, and revise hypotheses about the target language in order to mentally develop a new linguistic system. As such, the quantity and quality of input can directly mediate the mechanism, process, and outcome of acquisition (cf. Long 1996; Pica 2002). In terms of quantity, it would only seem a truism that the more input the better.[2] Researchers, especially from a usage-based perspective (see, e.g., Ellis 2002, 2006), have argued that learning is essentially a statistical process whereby the learner is more or less a statistician counting the tokens of a given element in the input, often, though not always, guided by frequency-made salience. It follows that the more the learner sees an element appear in the input, the greater the odds are that they will mentally register it (Ellis 2002; Ellis & Collins 2009). Yet, it has also been recognized (see, e.g., Gass & Mackey 2002) that second language (L2) learners can behave in defiance of frequency effects (Rast & Dommergues 2003). In spite of their high frequency, certain elements of input

[2] This assumption seems worth revisiting. We thank a reviewer for bringing the Endress and Bonatti (2007) study to our attention, which essentially showed that while we all possess the capacity to compute statistical properties of environmental stimuli, these computations were insufficient to extract structure. Structure-extracting mechanisms operate extremely rapidly; more exposure to input does not assist them. Thus, what may be needed for extracting structural information is not so much more exposure to the same inputs as exposure to different inputs that could reveal paradigmatic contrasts.

do not, if ever, become salient to the learner and, thus, evade their attention. The bottom line is that not all of the language data are utilized by the learner. As Gass (1988: 201) put it, "some language data pass through to the learner and some do not."

The phenomenon that input is only partially utilized by the learner was first touched on in print by Corder in his seminal work *The Significance of Learners' Errors* (1967), where he made a revolutionary distinction between *input* and *intake*: input is what is available to the learner, whereas intake is what goes in and is regulated by an internal mechanism – a built-in syllabus – in conjunction with the learner's current knowledge of the language. The impact of this conceptual split between input and intake has been so profound on SLA research (and for that matter, second language teaching) that on many levels it has altered its course, a discussion of which is beyond the scope of this chapter.

But, for one thing, since Corder (1967) researchers have not ceased wrestling with the relationship between input and intake, though exactly how input turns into intake has largely remained a mystery to date. While speculations abound, most of them appear to speak indirectly rather than directly, or opaquely rather than transparently, to the input–intake asymmetry. Consider Krashen's (1980, 1985) Input Hypothesis as a case in point. Krashen suggests that input needs to be comprehensible for it to become intake: "Humans acquire language in only one way – by understanding messages, or by receiving comprehensible input" (1985: 2). However, other than invoking the language acquisition device (LAD) as a mediator, the black box approach, Krashen provided no real pathway to understanding how comprehensible input ultimately becomes intake, nor did the ensuing empirical research designed either to find ways to make input comprehensible or to examine effects of input ostensibly made comprehensible on intake (see, e.g., Yano, Long & Ross 1994).

The need to treat intake as an object of inquiry cannot be overstated. Carroll (1999: 345) argues:

> Input is measurable in terms of objective properties of the signal – properties like signal frequency, duration and amplitude, spectral structure and so on. If intake is, in contrast, a selection of information from the signal (relevant for, e.g., the acquisition of word categories, morphological structure or relative clauses), then it can only be the output of some initial processing of the stimulus array. Intake is therefore not measurable in terms of objective properties of the stimulus array.

Precisely, it is this lack of objective properties as a learner-internal phenomenon that has prevented us from reaching a tangible understanding of how intake occurs. The conceptual and methodological challenge notwithstanding, progress has been made, and is manifest, primarily in theoretical terms, in the

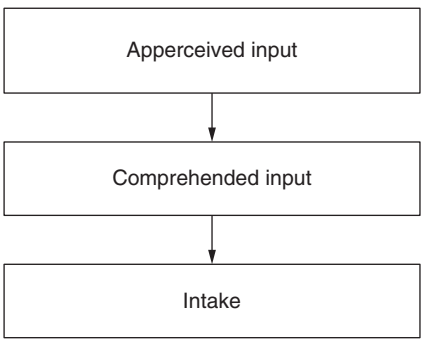

Figure 1.1 Gass's take on 'input to intake'

fact that several models of intake have been proposed since the 1980s, all attempting to explain why input does not equal intake and why the latter is selective (Boulouffe 1986; Carroll 1999; Chaudron 1985; Færch & Kasper 1980; Gass 1988; Sharwood Smith 1986; VanPatten 1996).

Models of input–intake

By way of illustration, Gass (1988, 1997), in sketching out a comprehensive framework of second language research, delved somewhat into the input and intake relationship by delimiting the input to "apperceived input" and "comprehended input" (see Figure 1.1). The implication is obvious: in order for input to become intake, it needs to have been noticed and understood. Invoking the term 'apperceived' from psychology, Gass underscored the status of prior knowledge as an anchor in the process of converting input to intake (cf. Boulouffe 1986):

> Apperception [is] thus the process of understanding by which newly observed qualities of an object are related to past experiences. In other words, past experiences relate to the selection of what we might call *noticed* (or apperceived) material. . . . apperception is not equivalent to perception. In perception an object (or in this case a linguistic form) is present in our senses, whereas apperception, being an internal cognitive act, identifies that form as being related to some prior knowledge which has been stored in our experience. We can think of apperception as a priming device which tells us which parameters to attend to in analyzing second language data. (1997: 201–202; emphasis in original)

Hence, a bridge was built in the input-to-intake process between the input, the learner, and the learner's prior knowledge, which includes, but is not limited to, knowledge of the native language, knowledge of other languages, existing knowledge of the target language, world knowledge, and language universals. The prior knowledge, Gass argued, serves as "activators of