A cognitive theory of second language acquisition (SLA) seeks to explicate the psychological mechanisms that underlie comprehension and production and the means by which that competence develops in the mind of the learner. The focus in this chapter is on recent research on second language sentence processing and the contribution this research can make to this endeavor. Sentence processing research seeks to identify how and when the various sources of information, syntax, semantics, context, and so on contribute to real-time processing outcomes in comprehension (Tanenhaus and Trueswell 1995) and production (Bock and Levelt 1994). Although second language sentence processing is still in its beginning, this research enterprise builds on a voluminous psycholinguistic literature on (first language) sentence processing. For reviews see Mitchell (1994), Tanenhaus and Trueswell (1995), and Harrington (forthcoming). Studies on real-time processes in second language comprehension have only recently started to appear (Juffs 1998b; Juffs and Harrington 1995, 1996). Although the investigation of on-line processing has only begun, there are a number of studies that have examined processing issues using off-line measures1 (Gass 1987; Harrington 1987; Kempe and MacWhinney 1998; Kilborn 1989; MacWhinney 1987; Myles 1995; Rounds and Kana-agy 1998; Sasaki 1994; Ying 1996).

The purpose of this chapter is to examine the emerging research literature in second language sentence processing. Two approaches to understanding real-time
sentence comprehension are contrasted. The \textit{syntax-based} approach, represented in the principle-based parsing research used in Juffs and Harrington (1995, 1996) and Juffs (1998a, 1998b), characterizes the comprehension process as the application of autonomous syntactic principles. These principles serve as the exclusive basis for initial parsing decisions, which are subsequently fed to interpretative processes that evaluate and, if necessary, revise the initial parse (Pritchett 1992). Semantics, frequency, and contextual information are assumed to play no role in initial parsing decisions.

In contrast, \textit{constraint-based} models of sentence processing assume that comprehension is the result of the interaction of multiple sources of knowledge, linguistic, pragmatic, contextual, and real-world. This information is represented in the mind in a distributed manner and serves as probabilistic constraints on interpretation. The constraint-based approach is most readily identified with connectionist modeling (Elman 1993), which is represented in the SLA literature in studies by Broeder and Plunkett (1994), Ellis and Schmidt (1998), Gass (1999), and Sokolik (1999), and Sokolik and Smith (1992).

Each approach is examined by first setting out the theoretical foundations and then describing how the approach investigates on-line sentence processing effects. The key questions that frame current research are identified and representative studies discussed. The strengths and limitations of the respective approaches are identified, and the chapter concludes with a discussion of the implications this line of research has for the development of a cognitive theory of SLA.

---

**SLA as a Cognitive Science**

This chapter approaches SLA as a cognitive science. Cognitive science seeks to understand the mental representations responsible for the higher-order mental functions (e.g., vision, language, categorization). Cognitive theory in SLA has been closely identified with the information-processing paradigm (McLaughlin and Heredia 1996). Information processing was the dominant metaphor for cognition in the 1960s and 1970s and remains so for many researchers today (Palmer and Kimchi 1986). Although the term is widely used, it is worthwhile to specify what the paradigm entails. An \textit{information process} is the means by which a system makes systematic responses to particular environmental conditions. The responses are typically goal oriented and serve an adaptive purpose for the system or the organism, which varies its \textit{behavior} in response to differing environmental conditions (Stillings et al. 1995). An understanding of an information process thus
includes an account of the information that the system needs to achieve its goals, as well as the means by which it uses that information. These means are usually described as subskills that contribute to the higher level information process, and the approach assumes that these component skills can be studied productively in isolation from the larger system. Because information processes take place in time, reaction time becomes a key variable in understanding the subskills involved. Finally, the information-processing approach assumes that the mind is a general purpose symbol processing system that is subject to capacity constraints. The information-processing paradigm has been and continues to be a highly productive source of SLA research (Bialystok 1994; DeKeyser 1995; Hatch 1983; Hulstijn 1990; Hulstijn and Hulstijn 1984; McLaughlin 1990; O’Malley and Chamot 1990; Robinson 1996; Schmidt 1992; Skehan 1991, 1998b; Van Patten and Cadierno 1993). For a more complete description of the information-processing paradigm in SLA, see McLaughlin and Heredia (1996).

Sentence processing research draws heavily on the information processing tradition for theory and research methodology. However, it also differs in three significant ways. First, the information-processing paradigm takes as a given the notion that the mind is a general purpose symbol processor, an assumption that is a major issue of contention in sentence processing research (Fodor and Pylyshyn 1988). The syntax-based and constraint-based accounts described in this chapter represent, respectively, classical and connectionist views of cognition and language (MacDonald and MacDonald 1995). The classical view sees the mind as a nonprobabilistic computational machine that carries out discrete operations on symbols (Fodor and Pylyshyn 1988). The connectionist perspective, in contrast, eschews a symbolic level of representation, characterizing language knowledge instead in distributed, probabilistic terms. Over the past decade, the debate over whether human cognition requires symbolic computation has been a—if not the—major issue in the cognitive science of language.

Second, as an account of SLA, information-processing approaches have traditionally left unspecified the nature of the linguistic knowledge that the learner acquires and uses (Newmeyer 1987). The nature of linguistic knowledge representations and how this knowledge is exploited in real-time comprehension is of central concern in sentence processing research and has profound implications for models of how this knowledge is developed. The third way in which sentence processing research differs from traditional information-processing research concerns the temporal dimension. Integral to any sentence processing model is an account of how comprehension processes are carried out in real time. The availability and exploitation of the various sources of information contributing to comprehension (e.g., linguistic, semantic, conceptual) under the pressure of online performance is a central issue among competing sentence processing approaches. In contrast, the focus in SLA information-processing research has been on identifying the conditions that affect learning outcomes, as in the implicit
versus explicit processing of input (Robinson 1997), or mechanisms, as in the need to attend to input (Schmidt 1990).

Although differences exist between sentence processing research and that in the information-processing tradition, the two approaches are complementary. Both will play an important role in the development of a cognitive theory of SLA. The remaining part of the chapter describes the contribution that L2 sentence processing research can make to our understanding of SLA in cognitive terms.

A Syntax-Based Approach to Sentence Processing

The syntax-based approach ascribes a central role to syntactic processes in sentence comprehension. In this section the theoretical foundation of the approach is described, and studies applying the framework to L2 processing are presented.

Language Processing as Symbol Computation

In the syntax-based processing approach, cognition and language are characterized as a symbol manipulation process (Newell, Rosenbloom, and Laird 1989). The symbolic approach assumes that knowledge is represented in the mind directly in symbols and that computations, specified in rules, are carried out on these representations. In natural language computation these symbols include phonemes, morphemes, grammar rules, and so on, and the processor works directly on these elements to yield an interpretation. The level of syntactic representation is assumed to be independent of the semantics of the specific items involved, in the same way that the computation of an algebraic equation (e.g., \( a + b = c \)) is the same, regardless of the specific values of \( a \) and \( b \).

The symbolic view has been the dominant approach to cognition, and this is reflected in the sentence processing literature. Sentence processing has traditionally been characterized as a process of symbolic computation, cast in the form of structure building (Mitchell 1994). The role of syntactic structure is thus of primary concern, and from the outset the interest has been in how the sentence processor (or parser) builds a syntactic structure that ultimately leads to an interpretation of the sentence (Frazier 1987). Fundamental insights into how this structure building proceeds have come from examining the processing of ambig-
uous language structures (e.g., visiting relatives), where structural alternatives are thrown into sharp relief. Ambiguity resolution processes provide a window on processes that are difficult to observe otherwise.

Syntax-Based Processing

Syntax-based approaches to processing ascribe a central role to syntactic knowledge in the sentence interpretation process. Syntactic knowledge consists of an autonomous competence grammar, and a principled distinction is made between the mechanisms responsible for lexical processing (e.g., word recognition and lexical access) and syntactic processing. Often referred to as two-stage models (Frazier and Fodor 1978), the syntactic parse is carried out rapidly using the minimal syntactic category information needed to complete the initial parse. The initial parse is then output to an interpretative mechanism that matches it against semantic, contextual, and real-world information, ultimately yielding an interpretation.

Syntax-based approaches can be classified into two categories: principle-based (Crocker 1994; Frazier 1989; Pritchett 1992) and referential (Altmann and Steedman 1988; Crain and Steedman 1985). The two differ principally as to when referential context information figures in processing outcomes. In the next section, the application to second language processing of one kind of principle-based parser, the generalized theta attachment model (Pritchett 1992), is discussed.

A Syntax-Based Second Language Processing

Juffs and Harrington (1995; 1996) used a principle-based model to investigate on-line reading processes by advanced ESL learners. The generalized theta attachment (GTA) model (Pritchett 1992) was used to assess the relative contribution of processing difficulty and (UG-based) grammatical knowledge to learner performance in on-line reading and grammaticality judgment tasks. Syntactic processing in the GTA model is driven by the assignment of thematic roles (e.g., agent, theme, goal) to elements in the input string. Lexical items can appear in more than one argument structure, and it is the possibility of multiple argument structure interpretations that can lead to processing ambiguities.

The GTA parser seeks to build as complete a structure as possible by assigning all thematic roles as soon as possible. As each word comes through the parser, syntactic principles (e.g., theta attachment, case, binding) are assigned so as to realize the most complete structure possible for the local string. (See Juffs and
Harrington 1995 and Juffs in this volume for details.) It is assumed that every NP must eventually be associated with a specific thematic role and that the parser selects the reading that imposes the lowest cost on the system. Processing difficulties thus arise as the result of unfulfilled thematic role assignments.

Juffs and Harrington (1995) were interested in the asymmetry evident in L2 learner performance on grammatical judgements for wh-structures, exemplified in (1) and (2).

(1) Who did Ann believe ___ likes her friend? (Subject extraction from finite clause)

(2) Who did Ann believe her friend likes ___ ? (Object extraction from finite clause)

The sentences differ structurally in that (1) involves the extraction of the pronoun who from the subject site (indicated by the gap), while (2) involves extraction from the object site. Earlier research showed that ESL learners experience more difficulty judging the acceptability of subject extraction sentences like (1) than their object extraction counterparts in (2) (Schachter and Yip 1990). Structures that involve constraints on wh-movement are of interest because they provide a testing ground for the putative effects of UG in adult SLA, given that these constraints are not assumed to operate in all languages (Epstein, Flynn, and Martohardjono 1996). Poor performance on the subject extraction sentences by adult L2 learners from languages in which wh-movement constraints are not assumed to operate (e.g., Chinese) have thus been used as evidence for lack of access to these principles by L2 learners (Schachter and Yip 1990). However, while L2 learners did have difficulty in the subject structures (1), they were sensitive to wh-constraints in judgments on other types of wh-structures, including the object extraction types in (2). This raised the possibility that the difficulties encountered by learners on these particular structures might be due to processing deficits rather than competence deficits. It was this possibility that Juffs and Harrington (1995) explored.

Reading times and grammaticality judgments of advanced Chinese ESL learners on sentences like (1) and (2) were collected in an on-line reading task. The key area of comparison in the study was processing times for the region immediately following the main verb believe. In the subject extraction sentences, this region follows the extraction site (the “postgap” region). Reading times were sharply higher in the postgap region for the subject extraction from clauses in (1), the slowdown in processing mirroring the observed decline in accuracy for these forms. Given that the Chinese learners in the study showed sensitivity to wh-movement constraints in general, the relative difficulty encountered on the subject
extraction sentences was attributed to greater processing demands that resulted from the realignment involved in the correct assignment of thematic roles in these structures. In sentence (i), the GTA model predicts that the parser will initially interpret the main verb believe as an NP complement and will posit a complete grammatical sentence with an object gap, as in Who does Ann believe? The appearance of the verb like then forces the object gap to be realigned as an embedded subject trace, Who does Ann believe likes? The authors thus concluded that the demands of the on-line realignment, and not the availability of the wh-movement constraints, may be responsible for the observed differences in performance on the respective structures evident in the earlier research (Schachter and Yip 1990).

The GTA model was also used by Juffs to study other types of on-line processing ambiguities that have received considerable attention in the L1 sentence processing literature, including so-called garden path structures, as Before Mary ate the pizza arrived from the restaurant. (Juffs 1998b), and ambiguities that arise from main verb/reduced relative clause readings of the verb, as in The leader defeated in the election versus The leader defeated the amendment... (Juffs 1998a). In both studies, the focus was on the role that L1 argument structure and UG constraints might play on L2 processing outcomes. Results from both sets of findings indicated that processing outcomes were affected by a range of factors that may involve, but are not limited to, UG-based crosslinguistic variation.

Syntax-Based Approaches in SLA

The principle-based parser used in the studies just discussed is representative of the syntax-based approach to on-line sentence processing. The studies also represent the first attempt to apply the framework to understanding second language processing issues. In this section the implications of these findings for a theory of second language processing and, in turn, to a cognitive theory of SLA are be considered.

The use of a competence-based grammar in the Juffs and Harrington (1995) study allows the interface of syntactic knowledge representations and processing mechanisms to be systematically examined. The grammar provides a formal, testable set of predictions that permits the researcher to isolate and identify the relative contribution of the grammar and the processor to real-time language comprehension. In adult L2 comprehension, the understanding of this interaction is made more complex by the wide range of individual differences in knowledge and processing across learners and by the potential effect of the first language on both these dimensions (Juffs 1998a).
However, the prominence accorded syntactic processes in the GTA model may also be problematic. In the studies cited, the processing variation evident in the word-by-word reading times, both within and across individuals, was interpreted in terms of structural complexity demands as predicted by a GB-based parsing model. However, many researchers find the assumptions made by these models to be overly restrictive or even misplaced (Tanenhaus and Trueswell 1995). Research has shown that a range of nonsyntactic information can neutralize or even overturn predicted structural processing effects (Altman and Steedman 1988). It has been demonstrated that the manipulation of prior discourse can bias responses toward or away from the main verb/reduced relative clause readings in a garden path sentence like *The waiter served calzone complained*. For example, the reduced relative clause reading can be elicited by prefacing the test sentence with *Two waiters were served different types of Italian food. The waiter served calzone* . . . (Trueswell and Tanenhaus 1994). Semantic effects can also affect initial parsing decisions. The fact that *waiter* in *The waiter served* . . . is animate makes an ambiguous reading more likely than would the presence of an inanimate noun compare *The spaghetti served* . . . (Ferriera and Clifton 1986; Trueswell, Tanenhaus, and Garnsey 1994).

The frequency with which particular verbs appear in particular argument structures can also bias sentence interpretations. The main verb interpretation of *served* in the example sentence is possible with both a transitive and intransitive argument structure, while the reduced relative clause reading requires a transitive argument structure. Verbs that appear more frequently in transitive argument structures are less likely to be given an intransitive reduced relative interpretation, despite the fact that both structural interpretations are available (Trueswell 1996). These findings indicate that sources of information excluded in the GTA account, like prior context and frequency biases, play an immediate and important role in on-line sentence interpretation.

Finally, the syntax-based approach provides little insight into how knowledge of the L2 develops. Syntax-based approaches all assume some form of competence grammar, with as yet little understanding of the mechanisms by which it develops (Gregg 1996).

In the Juffs and Harrington study, for example, the subjects were advanced L2 learners, and all were considered to possess a final-state L2 grammar. The effects observed thus reflected the outcome of the learners’ experience, and provide little insight into how that knowledge developed, or the course of development that took place.

In summary, syntax-based approaches like the GTA provide an explicit, testable set of processing claims. However, they also exclude factors that have been shown to play a central role in sentence processing. As a result, the contribution that this approach can make to a cognitive theory of SLA, where learning is the primary concern, remains uncertain.
A Constraint-Based Approach to Sentence Processing

Constraint-based models of sentence processing offer a sharp contrast to the syntax-based approach described in the preceding section. These models characterize sentence interpretation as a highly interactive process, in which syntactic, lexical, and semantic-conceptual information interact closely to constrain on-line comprehension (Tanenhaus and Trueswell 1995). In this section the theoretical foundation of the approach are described, and constraint-based models of L1 and L2 processing are examined.

Language as a Distributed Probabilistic System

There are several important features of the constraint-based approach that serve to distinguish it from the syntax-based approach. First, constraint-based models assume that language is represented as distributed, graded features, in contrast to the traditional symbolic view. Linguistic knowledge is represented in the mind as a complex pattern of associative links between units. These units, or microfeatures (Clark 1993), are smaller than the traditional units of linguistic analysis (e.g., phoneme, word, phrase, rule). The distributed knowledge representations produce rule-like behavior but do not assume the existence of explicit rules, which is the basic stuff of symbolic models of language and cognition. The effect of this experience is incremental, or graded, with patterns of activation corresponding to words, grammatical structures, and so on, undergoing constant modification as a result of experience. Repeated exposure strengthens the activation level of a representation, while decreasing exposure serves to weaken and, in certain cases, extinguish it. The graded nature of the representations contrasts with traditional symbolic approaches, in which these representations are assumed to be learned in an all-or-none manner (Hintzman 1993).7

The distributed, graded nature of knowledge representations also means that language knowledge is readily describable in probabilistic terms, another distinguishing feature of the approach (Seidenberg and McDonald 1999). Knowledge of a given form can range on a continuum from near certitude, where the structure is always used correctly, through an intermediate state, where appropriate usage varies, to a random state, where the incidence of correct use of the structure does not rise above chance. Probabilistic models have an advantage over rule-based models in capturing the variable nature of behavior, which is important in contexts where variation itself is of theoretical interest, as is the case in SLA. The
probabilistic, statistical nature of language processing is in sharp contrast with classical symbolic models where rule knowledge is represented and processed in a discrete fashion (Fodor and McLaughlin 1990).

The final, and potentially most significant, characteristic of the constraint-based perspective is the direct relationship it assumes between processing and learning. Novel input is processed on the basis of previously stored experience, with the act of processing itself changing the strength of existing knowledge representations. As a result, the mechanisms involved in processing input are also responsible for learning new knowledge (Sharkey 1996). Furthermore, the cognitive mechanisms responsible for the development of activation strengths are not specific to language learning but are shared by other higher order cognitive processes. As a theory of language learning, it characterizes language development as the result of the interaction between the learning environment and domain-general learning capacities of the individual, or what is termed an emergent property of the system (Ellis 1998).³

Sentence Processing as Constraint Satisfaction

Sentence interpretation in the constraint-based approach is described as an interactive process of constraint-satisfaction (McClelland, Rumelhart, and Hinot 1986).³ A good example is the lexicalist constraint-based model developed by MacDonald and her coworkers (MacDonald, Perlmuter, and Seidenberg 1994). In the model, units corresponding to the various information types are activated in parallel, with the strength of activation of a particular unit or set of units reflecting the type, number, and strength of the links it shares with other units in the system. Alternative structures are activated to differing degrees, and the interpretation depends on which alternative the system ultimately settles on. For example, the resolution of the temporarily ambiguous served in the example sentence The waiter served calzone (complained) involves structural decisions concerning argument structure, tense, and voice of the verb, which can be highly sensitive to frequency and context effects (Trueswell 1996).⁴

Constraint-Based Models of Sentence Processing:

Connectionist Sentence Processing

In contrast to the vast sentence processing literature in the symbolic tradition, connectionist accounts of sentence processing are fewer in number (Christiansen and Chater 1999). Connectionist approaches have been highly successful in char-
acterizing learning in local domains of processing phonological, lexical, and morphosyntactic level. Models have been trained to accurately assign verb tense (Plunkett and Marchman 1993) and to resolve lexical ambiguity (Kawamoto 1993) with a fairly simple mechanism that works on input, without recourse to prior assumptions as to the nature of the representations, such as the existence of a competence grammar. The ability of the model to learn these domains challenges the symbolic approach and provides a model that combines processing and learning in a single account.

Connectionist accounts of specific morphosyntactic domains in L2 development have also appeared, including the mapping of lexical items onto thematic roles (Gasser 1990), gender (Sokolik 1990; Sokolik and Smith 1992), verb morphology (Broeder and Plunkett 1994), and number (Ellis and Schmidt 1998). A representative application of connectionist theory and method to a domain of L2 learning is Ellis and Schmidt (1998). The focus of the study was on the frequency by regularity interaction that has been observed in the processing of past tense verbs in English (Prasada and Pinker 1993). English native speakers appear to produce the past tense of regular verbs (play-played) and irregular verbs (run-ran) differently. The time it takes to produce the past tense of an irregular verb is closely related to its frequency: High frequency past forms are produced quickly, and low frequency forms are much slower. Regular verbs, in contrast, do not appear to be sensitive to stem frequency, as high-frequency and low-frequency stems take approximately the same time to produce.\footnote{This pattern of results led Pinker and others to posit two separate mechanisms as being responsible for the production of the English past tense (Pinker and Prince 1994). Irregular verbs, which are sensitive to frequency effects, were posited to be stored as individual items in associative memory, with production a matter of retrieval from memory. Regular pasts, in contrast, are generated by a rule binding the past tense morpheme to the stem at the time of production. Beck (1997) presented similar results for advanced L2 learners for production on the regular verb forms but did not find a frequency effect for irregular verbs.}

Ellis and Schmidt (1998) examined how these forms were learned both by human participants and in a connectionist simulation. In the study, a group of subjects was taught the plural morphology of an artificial language, which preserved the frequency and regularity features of the past tense verb data in Prasada and Pinker (1993). The findings of the subsequent simulation studies also revealed a frequency by regularity interaction similar to the earlier study.\footnote{This led Ellis and Schmidt to conclude that a symbolic, rule-based approach was not needed to account for the results (Pinker and Prince 1994).\footnote{The Ellis and Schmidt study demonstrates how connectionist modeling can yield insight into the learning of morphosyntax. However, the capacity of connectionist models to capture the kind of empirical data relevant to key issues in sentence processing is still questionable (Carroll 1995). The challenge for these}}
models is to capture the complex structural relations that are integral to language but that are not readily apparent from simple co-occurrence statistics. These relations include long-distance dependencies among elements manifest in, for example, number agreement, embedding, and the filler-gap structures examined in Juffs and Harrington (1995). It has long been argued that these phenomena demand an autonomous level of syntactic representation that is represented naturally in symbolic approaches (Pinker and Prince 1988).

Early connectionist accounts of sentence-level syntax "hard-wired" the relevant linguistic structure in a distributed connectionist architecture (Small, Cotrell, and Shastri 1982). Thus, the models were not an alternative to symbolic accounts but provided a formal structure in which to implement them. Although these implementations provided insight into the interface between symbolic and connectionist accounts, a more radical goal for many connectionist researchers has been to provide an alternative to symbolic accounts of language. The attempt has been to develop models that would be able to learn linguistic structure from sequences of words, that is, without assuming an apriori competence grammar (Christiansen and Chater 1999). Two approaches have been taken.

Statistical connectionist models (Charniak 1993) extract syntactic structure by training a model on a corpus that contains sentences that are tagged for part of speech and then using it to assign the appropriate grammatical structure to novel sentences (Stolcke 1991). Although the model assumes only general learning algorithms, much of the task of extracting linguistic structure is completed in tagging the corpus on which the model is trained.

A more radical class of connectionist models attempts to learn syntactic structure from sequences of words that are not premodified. Elman (1993) used a simple recurrent network (SRN) model to predict the next word for sentences generated by a small context-free grammar that included key syntactic structures (e.g., subject-verb agreement, argument structure variation, and subject and object relative clauses). The model was able to acquire some of the regularities in the grammar, notably the ability to identify agreement relations across intervening words (Elman 1993). Elman's findings were extended by others, including Christiansen and Chater (1994), who included more complex grammatical structures, and Ellis (1997), who examined long-distance dependencies in an SLA context (Ellis and Schmidt 1997).

These findings marked an important advance for connectionist models of sentence processing but had a significant drawback. The models used a limited vocabulary and small ("toy") grammar, and great difficulty was encountered in trying to scale them up to account for the kinds of empirical data on ambiguity effects (Christiansen and Chater 1999). More recent attempts to fit on-line reading data on garden path (Tabor, Juliano, and Tanenhaus 1997) and center-embedded structures (Christiansen and Chater 1999) have had some success, but the research is still in its infancy. To date, no work has been done on second language sentence processing.
Constraint-Based Models and SLA Theory

The constraint-based approach has appeal for SLA research. The approach characterizes processing as the interaction of multiple sources of knowledge that develop in a graded, probabilistic fashion. As such, it can readily capture the variation across learners, languages, and settings that is an integral part of adult SLA. The connectionist models considered here posit a single, unified account for learning and processing and thus provides an explicit characterization of the developmental process. This is a pressing need for current cognitive SLA theory, which still knows little about the transition mechanisms responsible for the development of L2 knowledge (Gregg 1996).

However, connectionist processing accounts of key structural effects in real-time sentence processing have yet to appear. It is still uncertain whether these models can overcome the current limitations and establish themselves as legitimate alternatives to the symbolic syntax-based models of sentence processing. One alternative, not discussed here, is mixed models that combine the advantages of distributed representation with symbolic structural knowledge (Marcus 1998; Steedman 1999).

Implications of Sentence Processing Research for SLA Theory

In the final section of the chapter implications of sentence processing research for a cognitive theory of SLA are briefly discussed. Areas in which research on real-time language processing will make an increasingly important impact in SLA theory are in our understanding of the effects of the L1 on L2 use, on the mechanisms responsible for L2 development, and on SLA research methodology.

Cross-Linguistic Processing

Sentence processing research affords an important window both into the workings of the human speech processing mechanism and into the organization of the mind and behavior (Carpenter, Miyake, and Just 1995). A better understanding of the explicit mechanisms responsible for the use of a second language in real time will
bring more precision to the notion of processing, a term that is used in the current SLA literature with varying degrees of rigor (Sharwood Smith 1993).

The research discussed here also marks a move toward investigating possible L1-based processing effects on-line, which have not been studied in the past (e.g., Harrington 1987; Kilborn 1989; MacWhinney 1987; Sasaki 1994; Ying 1996). Recent works by Juffs (1998a, 1998b) raise important, and extremely complex, issues concerning the interaction of the L1 and the L2 in processing.

The Relationship between Processing and Learning

A fundamental difference between the two approaches discussed in this chapter concerns the relationship between processing and learning. The syntax-based approach does not directly address the issue, while the constraint-based models provide a parsimonious, if yet to be established, account of how knowledge of a second language develops. A better understanding of how and when the various sources of information are exploited in real-time will inform a model of SLA, regardless of which approach is taken.

Sentence Processing Research Methodology

The fact that SLA researchers have only started to examine sentence processing by second language learners is attributable to several factors. One limitation has been the lack of technical resources and methodological expertise required to carry out the research. Like all cognition, the processes involved in on-line sentence processing are only indirectly observable and unfold on a millisecond time scale. The technical, methodological, and, to some extent, conceptual tools needed to carry out this research have in the past not been readily available. However, that is changing with the appearance of usable and affordable data collection and analysis software tools.

Conceptually, the treatment of second language learning as a probabilistic process has major implications for research methodology and design, particularly in terms of the statistical tests used to establish reliable relationships between data. Null hypothesis testing, which has long been the statistical method of choice in experimental social science, assumes a discrete criterion value at which the null hypothesis is either rejected or retained and statistically “significant” difference inferred. Probabilistic processes, in contrast, are best described with correlational statistics, where the relative strengths (correlations) of multiple cue associations are of central theoretical interest (Howson and Urbach 1989). Sentence processing
wherein multiple cues naturally occur and interact is a prime example of multiple cue interaction. Correlational statistics have traditionally been considered problematic for establishing causal relations between variables, with a resulting (pronounced) bias toward null hypothesis testing. However, the need in sentence processing models to capture the interaction of multiple variables means that probabilistic statistical techniques, notably Bayesian modeling, will increasingly vie with the traditional two-valued statistical tests like the t-test or ANOVA.

CONCLUSION

The purpose of this chapter has been to examine the emerging research on second language sentence processing. Although still in its infancy, the field should make an increasingly important contribution to our understanding of the cognitive basis of second language learning and use. In addition, findings from research on second language processing will also provide further insight into the working of the human speech processing mechanism, in particular, and language and cognition, in general.

NOTES

1. Off-line measures include grammaticality judgments, categorization tasks, and true-false measures. They are off-line in the sense that the response, in the form of a judgment, categorization, and so on, is the result of the process. It provides only indirect evidence as to how the process was carried out in real time.

2. The thematic role of a phrase is the general semantic role that the phrase (called an argument) plays in relation to its predicate, with the various thematic roles possible for the phrase described by the argument structure of the predicate. Thematic roles (also called theta roles) include such entities as Theme (or Patient), which signals the entity that is undergoing the effect of some action, as Tom does in Tom fell over; Agent (or Causer), in which Tom is the instigator of an action, as in Tom killed Bob; and Goal, which is the entity toward which something moves, as in home in Tom went.


4. The Chinese ESL participants in the study performed at levels comparable to native speaker controls on grammatical extractions of objects (2) but were less accurate than the native speakers on grammatical extractions of subjects in (1) (Juffs and Harrington 1995: 501).
5. The GTA assumes that the interpretation of *Who did Ann believe ______ likes her friend?* requires a matrix object trace to be reanalyzed as an embedded subject trace. This process is assumed to consist of three steps: a change in theta role assigner from *believe* to *like*; a change in theta role from internal to external; and a change in Case assignment from accusative to nominative. In contrast, object wh-ex extractions involve only the reassignment of theta/Case assigner from matrix verb to embedded verb and are assumed to be easier to parse (Juffs and Harrington 1995: 492–493).

6. Particulars of the GTA account itself have also been challenged. The model adopts a fairly radical approach to how and when head information is used relative to other kinds of information, especially the stipulation that a node that identifies possible thematic roles cannot be projected before the head of the phrase is encountered. It appears that this yields different processing effects in head-initial (e.g., English) and in head-final (Japanese) languages (Karmelev and Mitchell 1999; Sturt and Crocker 1996). This, of course, is not a criticism of syntax-based accounts as a class.

7. It should be noted that recent symbolic models have also incorporated graded representations (Carpenter, Miyake, and Just 1995).

8. Learning outcomes reflect the individual’s experience with the environment. However, the constraint-based perspective is not a tabula rasa view of language learning, in which the learner brings an empty and unstructured set of cognitive capacities to the task. Rather, the point of departure in attempting to account for learning outcomes is that innate predispositions, such as those proposed in generative theory, make only a minimal contribution to the learning task. The emergent view sees development not in terms of either nurture or nature but as a rich and complex interaction between the two dimensions (Elman et al. 1996).

9. See Rumelhart and McClelland 1986, chapter 1, for a description of constraint satisfaction models of language and cognition. The terminology can be confusing. Pritchett’s 1992 model has also been called a constraint-satisfaction model, although the constraints that are satisfied are strictly defined in terms of grammatical principles (Crocker 1994: 249). This, however, is an exception, as constraint satisfaction models are usually associated with the interactive approach discussed here (Boland 1997; Clifton, Frazier, and Rayner 1994; Tanenhaus and Trueswell 1995).

10. The ambiguity encountered in structures like *The waiter served the calzone complained* is cast as a problem of lexical ambiguity (at three levels), rather than as a structural parsing or thematic role assignment problem. Three levels of lexical ambiguity that must be resolved in order to interpret *served* in *The waiter served . . .*. The argument-structure level alternates between transitive (in which both an agent and a theme thematic role can be assigned) and intransitive (in which only an agent role is possible); the tense morphology level between the past tense and past participle; and the voice level between active and passive (MacDonald, Perlmuter, and Seidenberg 1994).

The interdependencies between these levels is schematically represented in the pattern of connections. For example, the reduced relative clause reading (*The waiter who was served . . .*) requires that waiter be in the passive voice. This places a constraint on the other two levels, as only one possible alternative exists at each of the other two levels, namely the past-participle tense morphology and the transitive argument structure. The conjunction of the passive voice and the simple past tense is not allowed (*The waiter was flew by the pilot*), nor is the passive possible with intransitive argument structure (*The waiter was slept*).
11. Clahsen et al. observed a similar interaction for number inflection in German (Clahsen, Rothweiler, Woest, and Marcus 1993).

12. Singular word stems were the input units, and the plural prefixes were the output units. The output units were either the regular plural or irregular forms. The network was trained by presenting as input the singulars, which would then activate one of the plural output units. At the outset, this process was random, with a given input stem activating any of the plural output units. Each time a given input unit activated a particular output unit, the model would compare the activation of that mapping with the activation weight of the correct output unit, given that input unit. The backpropagation learning mechanism was used to calculate the difference between the mapping produced and the desired mapping and to then make an incremental adjustment in weights. As a result, the next time the stem input was presented, it was closer to the correct output unit in terms of activation strength. All the input-output mappings were thus trained separately over many blocks of trials. In addition, Ellis and Schmidt also had a singular stem that was used to test how well the network could generalize to novel items. They found that there was a small tendency for the untrained singular to activate the regular plural (e.g., as wags is given as the plural for wag).

13. This oversimplifies things somewhat, as there are other systematic differences between regular and irregular verbs that also led Pinker to adopt a rule-based model for the regular verbs (Pinker 1991).