To appear in: Anne Cutler (ed). *Psycholinguistic Interrelationships* (working title), Lawrence Erlbaum Assoc.

1 Rational models of comprehension: Addressing the performance paradox

Matthew W. Crocker
Saarland University

A fundamental goal of psycholinguistic research is to understand the architectures and mechanisms that underlie language comprehension. Such an account entails an understanding of the representation and organization of linguistic knowledge in the mind and a theory of how that knowledge is used dynamically to recover the interpretation of the utterances we encounter. While research in theoretical and computational linguistics has demonstrated the tremendous complexities of language understanding, our intuitive experience of language is rather different. For the most part people understand the utterances they encounter effortlessly and accurately. In constructing models of how people comprehend language, we are thus presented with what we dub the *performance paradox*: How is it that people understand language so effectively given such complexity and ambiguity?

In our pursuit and evaluation of new theories, we typically consider how well a particular model is able to *account* for observed results from the relevant range of controlled psycholinguistic experiments (empirical adequacy), and also the ability of the model to *explain* why the language comprehension system has the form and function it does (explanatory adequacy). Interestingly, research over the past twenty-five years has led to tremendous variety in proposals for parsing, disambiguation, and reanalysis mechanisms, many of which have been realized as computational models. However, while it is possible to classify models – e.g., according to whether they are modular, interactive, serial, parallel, or probabilistic – consensus at any concrete level has been largely

elusive.

We argue here for an alternative approach to developing and assessing theories and models of sentence comprehension, which offers the possibility of improving both empirical and explanatory adequacy, while also characterizing *kinds of models* at a more relevant and informative level than the architectural scheme noted above. In the following subsections, we emphasize the important fact that a model's coverage and behavior should not be limited to a few "interesting" construction types, but must also extend to realistically large and complex language fragments, and must account for why most processing is typically rapid and accurate, in addition to modeling pathological behaviors. We then argue that while the *algorithmic* description of a theory is essential to adequately assess its behavior and predictions, the theory of processing must also be stated at a more abstract level, e.g., Marr's *computational* level (Marr, 1982). In addressing these issues, we suggest that many of the ideas from *rational analysis* (Anderson, 1991) provide important insights and methods for the development, evaluation, and comparison of our models. In the subsequent section, we then discuss a number of existing models that can be viewed within a rational framework in order to more concretely exemplify our proposals.

Garden Paths versus Garden Variety

One great puzzle of human language comprehension is how easily people understand language despite its complexity and ambiguity, which we have termed the performance paradox. More puzzling is the fact that research in human sentence processing pays relatively little attention to this most fundamental and self-evident claim. In contrast, sentence processing research has focused largely on *pathological* phenomena: a relatively small proportion of ambiguities causing difficulty to the comprehension system. Examples include garden-path sentences, such as the well-known main verb/reduced-relative clause ambiguity initially noted by Bever (1970):

(1) The horse raced past the barn fell

In such sentences the verb *raced* is initially interpreted as the main verb, and only when the true main verb *fell* is reached can the reader determine that *raced past the barn* should actually have been interpreted as a reduced relative clause (cf., *The horse which was raced past the barn fell*). In this relatively extreme example, readers may not be able to recover the correct meaning at all, while other constructions may be

interpretable but result in some conscious or experimentally measurable difficulty.

The idea behind such research is to use information about parsing and interpretation preferences, combined with the factors that modulate them – such as frequency, context, and plausibility – to gain insight into the underlying comprehension system (see Crocker, 1999, for an overview). While this empirical research strategy might be seen as tacitly assuming rapid and accurate performance in general -- relying on pathologies only as a means for revealing where the "seams" are in the architecture of the language comprehension system – existing models of processing typically focus on accounting *only* for these pathologies. Furthermore, with few exceptions, existing models can be considered toy implementations at best, with lexical and syntactic coverage limited to what is necessary to model some subset of experimental data. Thus while such models may provide interesting and sophisticated accounts of familiar experimental findings, they provide no account of more general performance. Many theories have not been implemented at all, making it even more problematic to assess their general coverage and behavior.

Models à la Carte

Within the general area of computational psycholinguistics, a striking picture emerges when one compares the state of affairs in lexical processing with that in sentence processing. While there are relatively few models of lexical processing which are actively under consideration (see Norris, 1999), there exist numerous theories of sentence processing with relatively little consensus for any one in particular (Crocker, 1999; Townsend & Bever, 2001, chapter 4). The diverse range of models stems primarily from the compositional and recursive nature of sentence structure, combined with ambiguity at the lexical, syntactic and semantic levels of representation. The result is numerous dimensions of variation along which algorithms for parsing and interpretation might differ, including:

- *Linguistic knowledge:* What underlying linguistic representations, levels, interfaces, and structure-licensing principles are assumed? How is lexical knowledge organized and accessed?
- *Architectures:* To what extent is the comprehension system organized into modules? What are the temporal dynamics of information flow in modular and non-modular architectures?
- *Mechanisms:* What mechanisms are used to arrive at the interpretation of an utterance? Are representations constructed

serially, in parallel, or via competition? How does reanalysis take place?

However, while the formal and computational properties of language logically entail that a large number of processing models is possible, the space of models should be constrained by available empirical processing evidence. To some extent this has been achieved. Virtually all models, for example, share the property of strict incrementality. That is, the parsing mechanism integrates each word of an utterance into a connected, interpretable representation as the words are encountered (Frazier, 1979; Crocker, 1996). Beyond this, however, there is little agreement about even the most basic mechanisms of the language comprehension system.

Sentence processing research has long been preoccupied, for example, by the issue of whether the human language processor is fundamentally a restricted or unrestricted system, with various intermediate positions being proposed. Broadly, the restricted view holds that processing is served by informationally encapsulated modules, which construct only one interpretation (e.g., Frazier, 1979; Crocker, 1996). Unrestricted, or constraint-based, models on the other hand, assume that possible interpretations are considered in parallel, with all relevant information potentially being drawn upon to select among them (MacDonald, Pearlmutter & Seidenberg, 1994; McRae, Spivey-Knowlton & Tanenhaus, 1998).

However, while there exists a compelling body of empirical evidence demonstrating the rapid influence of plausibility (Pickering & Traxler, 1998) and visual information (Tanenhaus, Spivey-Knowlton, Eberhard & Sedivy, 1995; Knoeferle, Crocker, Scheepers & Pickering, in press) during comprehension, falsification of restricted processing architectures has not been possible. Furthermore, there is no direct empirical evidence supporting parallelism, i.e., that people simultaneously consider multiple interpretations for a temporarily ambiguous utterance as it unfolds.

Another area where mechanisms have proven difficult to distinguish empirically is reanalysis: when does the parser decide to abandon a particular analysis, and how does it proceed in finding an alternative? Consider the following example:

(2) *The Australian woman saw the famous doctor had been drinking.*

There is strong evidence that, for constructions such as this, people initially interpret the noun phrase *the famous doctor* as the direct object of

saw (e.g., Pickering, Traxler & Crocker, 2000), raising the question of how people recover the ultimately correct structure, in which that noun phrase becomes the subject of the complement clause. Sturt, Pickering and Crocker (1999) defend a representation preserving repair model for recovering from misanalysis (Sturt & Crocker, 1996), while Grodner, Gibson, Argaman, and Babyonyshev (2003) argue the same data can be accounted for using a destructive, re-parsing mechanism. Again, two apparently opposing models appear consistent with the same empirical findings.

Challenges

In summarizing the discussion above, we identify four key limitations, some or all of which affect most existing accounts of human sentence processing. We suggest these have contributed to both the lack of generality and comparability of our models, which has in turn stymied convergence within the field:

Limited scope: Models traditionally focus on some particular aspect of processing, emphasizing, for example, lexical ambiguity, structural attachment preferences, word order ambiguity, or reanalysis. Few proposals exist for a unified, implementable model of, e.g., lexical *and* structural processing *and* reanalysis. To the extent that such proposals do exist (e.g., Jurafsky, 1996; Vosse & Kempen, 2000), they are still typically so narrow in coverage that assessing general performance is difficult.

Model equivalence: Some models, while different in implementational detail, are virtually equivalent in terms of their behavior. For example, the symbolic model proposed by Sturt and Crocker (1996) overlaps substantially with Stevenson's (1994) hybrid connectionist model with regard to what structures are recovered during initial structure building and reanalysis. Indeed, even the Grodner *et al* (2003) account might be considered as *functionally equivalent*: even though the precise reanalysis mechanism is fundamentally different from that of Sturt and Crocker (1996) and Stevenson (1994), the "state" of the models is fundamentally identical as each word is processed.

Measure specificity: Models often vary with respect to the kind of experimental paradigms and observed measures they seek to account for. Models of processing load have relied primarily on self-paced reading data (Gibson, 1998; Hale, 2003), while theories of parsing rely on a variety of measures (e.g., first pass, regression path duration, and

total time) from eye-tracking during reading (e.g., Crocker, 1996; Frazier & Clifton, 1996). Some recent accounts are built upon the visual world paradigm, which monitors eye-movements in visual scenes during spoken comprehension (e.g., Tanenhaus *et al*, 1995; Knoeferle *et al*, in press), thus measuring attention, not processing complexity. Even more extremely, some models are based almost exclusively on neuroscientific measures, such as event-related potentials (Friederici, 2002: Schlesewsky & Bornkessel, to appear), placing little emphasis on accounting for existing behavioral data.

Weak linking hypotheses: Establishing the relationship between a model and empirical data demands a *linking hypothesis*, which maps the model's behavior to empirically observed measures. In explaining reading time data, for example, various models have assumed processing time is due to structural complexity (Frazier, 1985), backtracking (Abney, 1989; Crocker, 1996), non-determinism (Marcus, 1980), non-monotonicity (Sturt & Crocker, 1996), re-ranking of parallel alternatives (Jurafsky, 1996; Crocker & Brants, 2000), storage and integration cost (Gibson, 1998), the reduction of uncertainty (Hale, 2003), or competition (McRae *et al*, 1998). In addition, most models make only qualitative predictions as to the relative degree of difficulty. Those models which attempt more quantitative links with reading time data (McRae *et al*, 1998) fail to account for how structures are actually built (unlike the models outlined above), and are also highly fit to individual syntactic constructions.

TOWARDS RATIONAL MODELS

On the basis of discussion thus far, it should not be concluded that theories of sentence understanding posit particular processing architectures and implementations arbitrarily. In addition to linguistic assumptions, models are often heavily motivated and shaped by assumptions concerning cognitive limitations. Marcus (1980), Abney (1989), and Sturt and Crocker (1996) propose parsing architectures designed to minimize the computational complexity of backtracking. Some models argue that the sentence processor prefers less complex representations (Frazier, 1979), or assume other restrictions on working memory complexity. Other models restrict themselves by adopting a particular implementational platform, such as connectionist networks and stochastic architectures, as a way of incorporating cognitively-motivated mechanisms (e.g., Stevenson, 1994; Vosse & Kempen, 2000; Christiansen & Chater, 1999; Sturt, Costa, Lombardo & Frasconi, 2003).

Indeed it seems uncontroversial that human linguistic performance is to some extent shaped by such specific architectural properties and cognitive limitations. It is also true, however, that relatively little is known about the extent to which this is the case, let alone the precise manner in which such limitations affect human language understanding. We therefore suggest that by focusing on specific processing architectures and mechanisms and cognitive limitation, theories of sentence processing are forced into making stipulations without concrete empirical justification, but which nonetheless impact upon the overall behavior of models.

An alternative approach to developing a theory of sentence processing is to shift our emphasis away from particular mechanisms, and towards the nature of the sentence processing *task* itself:

> *An algorithm is likely understood more readily by understanding the nature of the problem being solved than by examining the mechanism (and the hardware) in which it is solved.* (Marr, 1982, p.27)

The critical insight here is that it can be helpful to have a clear statement of what the goal of a particular system is – and the function it seeks to compute – in addition to a model of how that goal is achieved, or how that function is actually implemented. For example, a systematic preference for argument attachment over modifier attachment, as argued for extensively by Pritchett (1992), can be viewed as providing an overarching explanation for a number of different preference strategies in the literature. Indeed, Crocker (1996) argues that Pritchett's theory itself, which seeks to maximize satisfaction of syntactic and semantic constraints, can be viewed as realizing an even more general goal of human language processing:

> *Principle of Incremental Comprehension* (PIC): The sentence processor operates in such a way as to maximize comprehension of the sentence at each stage of processing. (Crocker, 1996, p.106)

Such a statement in itself says little about the specific mechanisms involved and is indeed consistent with a range of proposals in the literature. It is, rather, intended as a claim about what kinds of models can be considered, and a general explanation for why they are as they are (namely, because they satisfy the PIC). This claim goes beyond saying that comprehension is incremental, something that is true of virtually all current models, and predicts that at points of ambiguity, the preferred structure should be the one that is maximally interpretable: e.g., it establishes the most dependencies, or maximizes role assignment and reception.

Focusing on the nature of the problem thus shifts our attention to the goals of the system under investigation, and the relevant properties of the environment. Anderson (1991) notes that there is a long tradition of attempting to understand cognition as *rational*: not because it follows some set of normative rules, but because it is optimally adapted to its task and environment. On the assumption that the comprehension system is rational, we can derive the optimal function for that system from a specification of the goals and the environment. The *Principle of Incremental Comprehension* does this rather implicitly: it assumes the goal is to correctly understand the utterance, and the environment is one in which language is both ambiguous and encountered incrementally.

In order to determine more precisely the function that comprehension seeks to optimize, we need also consider computational constraints in order to avoid deriving a function that is cognitively implausible in some respects (e.g., construction and evaluation of all – possibly infinite – interpretations, seems relatively implausible). However, an important aim of this kind of analysis is to see how much can be explained by avoiding appeal to such constraints except when they are extremely well motivated.

It should be clear that in adopting a Marrian/Andersonian approach, we address several of the potential pitfalls that have plagued model builders to date: emphasis on what function is computed (Marr's computational level), rather than specific algorithms and implementations should lead to better consensus, and more straightforward identification of models which are equivalent (in that they implement the same function). Furthermore, the approach emphasizes general behavior and performance, rather than the construction of models that are over-fitted to a few phenomena.

Inspired by Anderson's rational analysis, Chater, Crocker and Pickering (1998) motivate the use of probabilistic frameworks for characterizing and deriving mathematical models of human parsing and reanalysis. Probabilistic models of language processing typically optimize for the likelihood of ultimately obtaining the correct analysis for an utterance (Manning & Schütze, 1999).[1]

---

[1] We can formally express the *Principle of Likelihood* (PL) using notation standardly used in statistical language processing (Manning & Schütze, 1999):

(eq 2) $$\hat{t} = \underset{t \in T : yield(t) = s}{\arg\max} P(t \mid s, K)$$

The expression simply states that, from the set of all interpretations T which have as their yield the sentence s, we select the interpretation t which has the

This goal of adopting the most likely analysis, or interpretation, of an utterance seems plausible as a first hypothesis for a *rational* comprehension system. That is, in selecting among possible interpretations for an utterance, adopting the most likely one would be an optimally adaptive solution. Given our overriding assumption of incremental processing, this selection can also be applied at each point in processing: prefer the (partial) interpretation that is most likely, given the words of the sentence that have been encountered thus far.

There are some very important and subtle issues concerning our use of probabilities here. Firstly, using a probabilistic framework to reason about, or characterize, the behavior of a system does not explicitly entail that people actually use probabilistic mechanisms (e.g., frequencies) but rather that such a framework can provide a good characterization of the system's behavior. That is, non-probabilistic systems could exhibit the behavior characterized by the probabilistic theory. Of course, (some) statistical mechanisms will also be consistent with the behavior dictated by the probabilistic meta-theory, but these will require independent empirical justification.

Furthermore, probabilities may be used as an abstraction. For example if a sentence s is globally ambiguous, having two possible structures, we might suggest that the probabilities, $P(t_1|s,K)$ and $P(t_2|s,K)$, for the two structures provide a good *estimate* or characterization of which is "more likely". This is a perfectly coherent statement, even though the real reason one structure is preferred is presumably due to a complex array of lexical and syntactic biases, semantics and plausibility, pragmatics and context (some or all of which may in turn be probabilistic). That is, we are simply using probabilities as a short-hand representation, or an abstraction, of more complex preferences, which allows us to reason about the behavior of the language processing system (see Chater *et al.*, 1998, for detailed discussion).

It is in general not possible to determine probabilities precisely, rather we typically attempt to *estimate* probabilities using frequency counts from large corpora or norming studies (McRae *et al*. 1998; Pickering *et al.* 2000). Indeed, the usefulness of likelihood models in computational linguistics has led to a tremendous amount of research into how probabilistic language models can be developed on the basis of data-intensive, corpus techniques (see Manning & Schütze, 1999, for both an introduction and survey of recent models).

---

greatest probability of being correct given the s, and our knowledge K.

In the following two sections we outline several examples of how the *Principle of Likelihood* has been applied to the development of particular models of language processing. Such models can be considered theories at Marr's algorithmic level, in that they provide a characterization of *how* the language processor implements the maximum likelihood function.

Lexical Ambiguity Resolution

Corley and Crocker (2000) present a broad-coverage model of lexical category disambiguation based on the *Principle of Likelihood*. Specifically, they suggest that for a sentence consisting of words $w_0 \ldots w_n$, the sentence processor adopts the most likely part-of-speech sequence $t_0 \ldots t_n$. More specifically, their model exploits two simple probabilities: *(i)* the conditional probability of word w_i given a particular part of speech t_i, and (ii) the probability of t_i given the previous part of speech t_{i-1}.[2] As each word of the sentence is encountered, the system assigns it that part-of-speech t_i which maximizes the product of these two probabilities. This model capitalizes on the insight that many syntactic ambiguities have a lexical basis (MacDonald *et al,* 1994), as in (3):

(3) *The warehouse prices/makes are cheaper than the rest.*

These sentences are temporarily ambiguous between a reading in which *prices* or *makes* is the main verb or part of a compound noun. After being trained on a large corpus, the model predicts the most likely part of speech for *prices*, correctly accounting for the fact that people understand *prices* as a noun, but *makes* as a verb (see Crocker and

---

[2] Formally, we can write this as a function which selects that part-of-speech sequence which results in the highest probability:

(eq 2) $$\hat{t}_0 \ldots \hat{t}_n = \arg\max_{t_0 \ldots t_n} P(t_0 \ldots t_n, w_0 \ldots w_n)$$

Directly implementing such a model presents cognitive and computational challenges. On the one hand, the above equation fails to take into account the incremental nature of processing (i.e. it assumes all words are available simultaneously), while on the other hand, the accurate estimation of such probabilities is computationally intractable due to data sparseness. Their approach, therefore, is to approximate this function using a bi-gram model, which incrementally computes the probability for a string of words as follows:

(eq 3) $$P(t_0 \ldots t_n, w_0 \ldots w_n) \cong \prod_{i=1}^{n} P(w_i \mid t_i) P(t_i \mid t_{i-1})$$

Corley (2002), and references cited therein). Not only does the model account for a range of disambiguation preferences rooted in lexical category ambiguity, it also explains why, in general, people are highly accurate in resolving such ambiguities.

Corley and Crocker's model provides a clear example of how we can use probabilistic frameworks to characterize both the function to be computed according to the rational analysis, and also to derive a practical, cognitively plausible *approximation* of this function which serves as the actual model (refer to (eq 2) and (eq 3) in footnote 2). Of course, subsequent empirical research might suggest the bi-gram model is inadequate and should be replaced by, e.g., a tri-gram model. Any such evidence, however, would only involve revision at the *algorithm level*, not of the overarching *rational analysis,* or *computational level*, since the tri-gram model still approximates the maximum likelihood function posited by the *Principle of Likelihood*.

Syntactic Processing

While it provides a simple example of rational analysis, Corley and Crocker's model cannot be considered a model of sentence processing, as it only deals with lexical category disambiguation. As noted above, directly estimating the desired probability of syntactic trees is problematic, since many have never occurred before. Thus, rather than trying to associate probabilities with entire trees, statistical models of syntactic processing typically associate a symbolic component that generates linguistic structures with a probabilistic component that assigns probabilities to these structures. A probabilistic context free grammars (PCFG), for example, associates probabilities with each rule in the grammar, and compute the probability of a particular tree by simply multiplying the probabilities of the rules used in its derivation (Manning & Schütze, 1999, chapter 11).

In developing a model of human lexical and syntactic processing, Jurafsky (1996) further suggests using Bayes' Rule to combine structural probabilities generated by a probabilistic context free grammar with other probabilistic information, such as subcategorization preferences for individual verbs. The model therefore integrates multiple sources of experience into a single, mathematically well-founded framework. In addition, the model uses a beam search to limit the amount of parallelism required.

Jurafsky's model is able to account for a range of parsing preferences reported in the psycholinguistic literature. However, it might be criticized for its limited coverage, i.e., for the fact that it uses only a small lexicon and grammar, manually designed to account for a handful

of example sentences. In the computational linguistic literature, on the other hand, broad coverage probabilistic parsers are available that compute a syntactic structure for arbitrary corpus sentences with generally high accuracy. This suggests there is hope for constructing psycholinguistic models with similar coverage, potentially explaining more general human linguistic performance. Indeed, more recent work on human syntactic processing has investigated the use of PCFGs in wide coverage models of incremental sentence processing (Crocker & Brants, 2000). Their research demonstrates that even when such models are trained on large corpora, they are indeed still able to account not only for a range of human disambiguation behavior, but also exhibit good performance on natural text. Related work also demonstrates that such broad coverage probabilistic models maintain high overall accuracy even under the strict memory and incremental processing restrictions (Brants & Crocker, 2000) that seem necessary for cognitive plausibility. Finally, Hale (2003) extends the use statistical parsing models to providing a possible explanation of processing load, rather than ambiguity resolution.

The Informativity Model

The models outlined above all begin with the assumption that the *Principle of Likelihood* best characterizes the function of the sentence comprehension system. It is important to note, however, that alternative rational analyses may emerge, depending on the precise definition of the problem. Chater *et al.* (1998) argue that a more plausible rational analysis of human sentence processing must take into account a number of important cognitive factors before an appropriate optimal function can be derived. In particular, they consider the following:

- Linguistic input contains substantial local ambiguity, which is resolved incrementally.
- People consciously consider only one preferred, or *foregrounded*, interpretation of an utterance at any given time during parsing.
- Immediate reanalysis is typically much easier than delayed reanalysis, and therefore is a lower cost operation.

In deriving a rational analysis of interpretation, Chater *et al.* argue that the human parser is optimized so as to incrementally resolve each local ambiguity as it is encountered (Church & Patil, 1982). The result of the analysis is a function which includes not only likelihood, but also another measure, *specificity*, which determines the extent to which a

particular analysis is "testable". That is, specificity measures the extent to which subsequent input will assist in either confirming or rejecting the foregrounded structure. On this account, the initially favored analysis is the one that is both "fairly likely" and "fairly testable". The measure, which they term *informativity* (I), balances *likelihood* (P) and *specificity* (S), such that the interpretation which maximizes the product of these two is foregrounded at each point in processing.[3]

This model contrasts with pure likelihood accounts in predicting that the sentence processor will prefer the construction of testable analyses over non-testable ones, except where the testable analysis is highly unlikely. The result will be a greater number of easy misanalyses (induced by less probable but more testable analyses), and a smaller number of difficult misanalyses (induced by more probable but less testable analyses). This in turn means that the ultimately correct analysis will usually be obtained quickly, either initially or after rapid reanalysis.

The most compelling empirical support for the *Principle of Informativity* stems from experiments by Pickering *et al.* (2000), in which the plausibility of a low frequency structural alternative (the NP-complement subcategorization frame for a verb like *realised*) was manipulated, as in *The athlete realized his* {*goals* vs. *shoes*} *... were out of reach*. Assuming a likelihood-based model, which would foreground an S-complement, there should be no effect of plausibility given that the low probability NP-complement option would no be considered during initial analysis.[4] Reading time experiments demonstrated, however, a striking asymmetry between frequency bias and actual processing performance, indicating that the low frequency alternative was immediately considered during on-line sentence comprehension. Pickering *et al*. argued that the low frequency NP-complement analysis is locally more 'specific', and hence can be evaluated earlier than the high frequency S-complement alternative. For a system with limited processing resources, such a strategy is advantageous, as it minimizes the cost of reanalysis.

Pickering *et al.* (2000) define the specificity of an analysis as a

---

[3] Again, we can formalize this straightformwardly as follows:

(eq 5) $$\hat{t} = \underset{t \in T : yield(t) = s}{\arg\max} \; I(t) = P(t) \bullet S(t)$$

[4] Though see Crocker & Brants (2000) for an explanation of why their model does in fact account for this data.

measure of how strongly that analysis constrains the sentence's continuation. A highly specific analysis entails that the parser has strong expections about the subsequent input. If these expectations are fulfilled, then this is taken as further support for the analysis, and parsing continues. If expectations are not fulfilled, the parser knows to immediately pursue an alternative analysis. Thus, Informativity predicts that the parser may prefer an analysis that is less probable than another, if it is more specific. While this leads to more misanalyses than a pure likelihood model, they are precisely those misanalyses from which the parser can recover quickly: an analysis that is potentially incorrect (i.e., improbable) would only be adopted if highly specific, hence the parser will be able to recognize and correct the error quickly.

As noted by Pickering *et al.* (2000), the *Principle of Informativity* differs crucially from the *Principle of Likelihood* in that it favors the construction of interpretable dependencies, thus providing an overarching rational analysis explanation for previously proposed strategies in the literature, such as Minimal Attachment (Frazier, 1979), theta-attachment (Pritchett, 1992), and the Principle of Incremental Comprehension (Crocker, 1996) among others.

The main point here, however, is not to argue whether the *Principles of Likelihood* or *Informativity* provide a better characterization of the function computed, but rather to highlight how different rational analyses can be developed, and their predictions, tested. Settling on a theory or analysis at Marr's computational level enables us to constrain and compare the models which approximate such a theory. Furthermore, it allows us to distinguish data which falsifies a particular model from data which falsifies the more general theory. This is crucial, since models will typically be an imperfect approximation of the theory (taking into account, e.g., cognitive limitations on memory or processing, or simple practical/implementational constraints), and hence a particular model may well make slightly differing predictions from the computational theory.

CONCLUSIONS

This chapter has argued for a shift in how we go about developing models of human language comprehension. We suggest that by adopting insights from rational analysis, we will not only make more progress in developing our *theories*, but also in building, evaluating and comparing our *models*.

1. Rational theories include a high-level characterization of the function

computed by the comprehension system, independent of specific architectural and mechanistic assumptions or stipulations. As such, a rational analysis provides both a predictive and explanatory basis for the mechanisms that implement it.

2. The existence of a rational theory can help in identifying models that are functionally similar, differing primarily in implementation, and hopefully assist in identifying points of convergence among theories.
3. Rational analyses derive from the primary observation that the comprehension is optimally adapted to the task of understanding. This places increased emphasis on explaining general performance, rather than modeling a handful of ambiguous constructions.

We have briefly summarized a collection of models that can be straightforwardly viewed as rational. Many probabilistic models of comprehension can be seen as deriving from the more general *Principle of Likelihood* (see also Jurafsky, (2003) for an overview). We have shown, however, that differing assumptions concerning the nature of the comprehension task can result in optimal functions other than likelihood, as in the case of the *Principle of Informativity*, and also observed that such an analysis provides greater compatibility with existing, non-probabilistic, proposals in the literature. Indeed, it is important not to conflate, *a priori*, probabilistic models with frequency-based models. While many researchers do assume that the probabilities in their models are derived from frequency of occurrence, we may also use it simply as short-hand for likelihoods which are derived from other sources (e.g., plausibility, rather then probability).

There are at least two weaknesses of the rational analysis approach. First, the relatively abstract nature of a computational theory results in a relatively weak linking hypothesis. Typically, the theory will provide only qualitative predictions about processing, e.g., which interpretation should be preferred. This is simply due to the fact that more precise accounting of observed measures, such as reading times, will be dominated by the specific mechanisms that implement the theory, and those of the other perceptual systems involved. For example, most of the variance in reading times is accounted for by factors such as word length and frequency (Keller, 2003). This "weakness" can actually be viewed positively, in that it allows us to distinguish the qualitative predictions of the theory from the more quantitative predictions of specific models which we may be considering as implementations of the theory.

Secondly, the approach is most appropriate in theorizing about cognitive systems that can be viewed as optimally adapted to their task

and environment. If the function of the system is shaped primarily by cognitive limitations or specific properties of the neural hardware, then such an analysis is seriously compromised. This contrasts starkly with the many models of sentence processing that are motivated precisely on the basis of cognitive limitations (working memory, parsing complexity) or specific processing architectures (e.g., connectionist networks, or modular information processing).

We argue here, however, that there is sufficient evidence for the adaptive nature of human comprehension – including the rapid use of frequency information, visual and linguistic context, plausibility and world knowledge, as well as more general evidence for the speed, accuracy, and robustness of the comprehension system – to warrant the pursuit of rational accounts.

Acknowledgements

The author would like to acknowledge the financial support of the DFG funded project ALPHA, (SFB-378: "Resource Adaptive Cognitive Processes"). This chapter has also benefited substantially from the comments and discussion received from participants of the MPI Four Corners Workshop series in Nijmegen, notably Harald Baayen and Anne Cutler, as well as the ongoing intellectual contributions from Nick Chater and Martin Pickering concerning many of the ideas presented here. Finally I would also like to thank my colleagues Pia Knoeferle and Marshall Mayberry for comments on a previous draft.

References

Abney, S. (1989). A computational model of human parsing. *Journal of Psycholinguistic Research, 18(1)*, 129-144.

Anderson, J.R. (1991). Is human cognition adaptive? *Behavioral and Brain Sciences*, 14, 471-517.

Bever, T. (1970). The cognitive basis for linguistic structures. In J. Hayes (Ed.), *Cognition and the development of language*. New York: Wiley. 279-362.

Brants, T. & Crocker, M.W. (2000). Probabilistic Parsing and Psychological Plausibility, In *Proceeding of the International Conference on Computational Linguistics (COLING 2000)*, Saarbrücken, Germany, 111-117.

Chater, N., Crocker, M.W., & Pickering, M. (1998). The Rational Analysis of Inquiry: The Case for Parsing. In N. Chater & M. Oaksford (eds), *Rational Analysis of Cognition*, Oxford, UK: Oxford University Press, 441-468.

Christiansen, M. H., & Chater, N. (1999). Toward a connectionist model of recursion in human linguistic performance. *Cognitive Science*, 23(2), 157-205.

Church, K. & Patil, R. (1982). Coping with syntactic ambiguity or how to put the block in the box on the table. *American Journal of Computational Linguistics*,

8(3–4), 139-149.
Corley, S. & Crocker, M. (2000). The Modular Statistical Hypothesis: Exploring Lexical Category Ambiguity. In M. Crocker, M. Pickering & C. Clifton, Jr. (eds.) *Architectures and Mechanisms for Language Processing.* Cambridge, UK: Cambridge University Press, 135-160.
Crocker, M. (1999). Mechanisms for Sentence Processing. In S. Garrod & M. Pickering (eds), *Language Processing*, London, UK: Psychology Press, 191-232.
Crocker, M. (1996). *Computational Psycholinguistics: An Interdisciplinary Approach to the Study of Language*, Dordrecht, NL: Kluwer.
Crocker, M. & Brants, T (2000). Wide Coverage Probabilistic Sentence Processing. *Journal of Psycholinguistic Research*; 29(6), 647-669.
Crocker, M. & Corley, S. (2002). Modular Architectures and Statistical Mechanisms: The Case from Lexical Category Disambiguation. In P. Merlo & S. Stevenson (eds), *The Lexical Basis of Sentence Processing.* Amsterdam: John Benjamins, 157-180.
Frazier, L. (1979). *On comprehending sentences: Syntactic parsing strategies*. PhD Thesis, University of Connecticut, CT.
Frazier, L. (1985). Syntactic Complexity. In D. Dowty, L. Kartunnen & A. Zwicky (eds), *Natural Language Parsing*, Cambridge UK: Cambridge University Press, 129-189.
Frazier, L. & C. Clifton, Jr. (1996). *Construal*. Cambridge, MA: MIT Press.
Friederici, A. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences*, 6, 78-84.
Gibson, E. A. F. (1998). Linguistic complexity: locality of syntactic dependencies. *Cognition*, 68 (1), 1–76.
Grodner, D, E. Gibson, E., Argaman, V. & Babyonyshev, M. (2003). Against repair-based reanalysis in sentence comprehension. *Journal of Psycholinguistic Research, 32(2)*, 141-166.
Hale, J. (2003). The information conveyed by words in sentences. *Journal of Psycholinguistic Research*, 32(2), 101-124,
Jurafsky, D.A (1996). Probabilistic Model of Lexical and Syntactic Access and Disambiguation, *Cognitive Science*, 20:137-194.
Jurafsky, D.A. (2003). Probabilistic modeling in psycholinguistics: Linguistic comprehension and production. In R. Bod, J. Hay & S. Jannedy (eds.), *Probabilistic Linguistics*. Cambridge, MA: MIT Press.
Keller, F. (2003). A probabilistic parser as a model of global processing difficulty. In proceedings of: *The 25th Annual Conference of the Cognitive Science Society*, Mahawah, NJ: Erlbaum, 646-651.
Knoeferle, P., Crocker, M., Scheepers, C. & Pickering, M. (in press), The influence of the immediate visual context on incremental thematic role-assignment: evidence from eye-movements in depicted events. *Cognition.*
MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, 101, 676-703.
McRae, K., Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1998). Modelling the influence of thematic fit (and other constraints) in on-line sentence

comprehension. *Journal of Memory and Language*, 38:283-312.

Manning, C. & Schütze, H. (1999). *Foundations of Statistical Natural Language Processing*, Cambridge, MA: MIT Press.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information.* San Francisco: W H Freeman.

Marcus, M. P. (1980). *A Theory of Syntactic Recognition for Natural Language*. Cambridge, MA: MIT Press.

Norris, D. (1999). Computational Psycholinguistics. In R. A. Wilson & F. C. Keil (eds.), *The MIT Encylopedia of Cognitive Science*. Cambridge, MA: MIT Press.

Pickering, M. & M. Traxler (1998). Plausibility and recovery from garden paths: An eye-tracking study. *Journal of Experimental Psychology: Learning, Memory and Cognition, 24*, 940-961.

Pickering, M. Traxler, M. & Crocker, M.W. (2000). Ambiguity resolution in sentence processing: Evidence against likelihood. *Journal of Memory and Language*; 43(3):447-475.

Pritchett, B. (1992). *Grammatical competence and parsing performance.* Chicago: University of Chicago Press.

Schlesewsky, M. & I. Bornkessel (to appear). On incremental interpretation: Degrees of meaning accessed during language comprehension. *Lingua.*

Stevenson, S. (1994). Competition and recency in a hybrid network model of syntactic disambiguation. *Journal of Psycholinguistic Research*, 23(4), 295-322.

Sturt, P., & Crocker, M. (1996). Monotonic syntactic processing: A cross-linguistic study of attachment and reanalysis. *Language and Cognitive Processes*, 11, 449-494.

Sturt, P., Pickering, M., & Crocker, M.W. (1999). Structural Change and Reanalysis Difficulty in Language Comprehension. *Journal of Memory and Language, 40(1),* 136-150.

Sturt, P., Costa, F., Lombardo, V., and Frasconi, P. (2003). Learning First-pass structural attachment preferences with dynamic grammars and recursive neural networks. *Cognition*, 88, 133-169.

Tanenhaus, M.K., Spivey-Knowlton, M.J., Eberhard, K.M. & Sedivy, J.E. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*, 632-634.

Townsend, D. J. & Bever, T. G. (2001). *Sentence comprehension: the integration of habits and rules.* Cambridge, MA: MIT Press.

Vosse, T., & Kempen, G. (2000). Syntactic structure assembly in human parsing: a computational model based on competitive inhibition and a lexicalist grammar. *Cognition*, 75, 105-143.