

Corpus and the Nature of Grammar Revisited*

One-Soon Her

I-Ping Wan

National Chengchi University

This paper is a response to Shei (2004), where the author makes two main claims: (1) the generative grammar is implausible based on evidence from psycholinguistic and biological studies, and (2) lexicon may prove to outweigh syntax and corpus linguistics is thus a more viable alternative. By clarifying some of the generativist views and the current state of affairs and also pointing out some of the inconsistencies in Shei's arguments, we aim to defend the generative paradigm as a worthy scientific pursuit and, more importantly, also to demonstrate that the generativist approach does not necessarily conflict with the functionalist approach in general or the corpus-based methodologies in particular. We maintain that both approaches may be necessary for a comprehensive view of *language* and *languages*.

Key words: generative grammar, functional grammar, psycholinguistics, corpus linguistics, innateness, autonomous syntax

1. Introduction

'Corpus and Grammar: What it isn't', by Chris Shei, was the most prominent article in the June 2004 issue of *Concentric: Studies in Linguistics* (30.1:1-18), an issue dedicated to the theme of corpus and grammar. The article's main claim is that evidence from psycholinguistic and biological studies disfavors the generative approach to grammar. Based on that claim, Shei urges the abandonment of this research paradigm and enthusiastically endorses corpus linguistics as a viable alternative. What prompted this reply is not so much this anti-generativist view *per se*. After all, such a view is not uncommon in the functionalist literature, though by no means shared by all, or even most, functionalist linguists, corpus linguists, computational linguists, or applied linguists. The fact is, irrespective of Shei (2004) and this reply, the functionalist versus generativist debate will no doubt go on in the foreseeable future. The reason we felt compelled to compose this reply is because the readers that Shei aims to sway are the more impressionable linguistics students rather than the more experienced professionals:

* The paper is due to the collaboration of the two authors, who share the views expressed here. However, the first author was primarily responsible for Sections 1-2 and 5-7, and the second author wrote most of Sections 3 and 4. Their respective contribution to the paper is 2/3 for the first author and 1/3 for the second author. We are very grateful for the comments and suggestions made by the two anonymous reviewers. However, we are solely responsible for the content of the paper. We also acknowledge the financial support provided by the following National Science Council grants: NSC-94-2411-H-004-030 and NSC-95-2411-H-004-027 to O-S Her and NSC-94-2411-H-004-001 and NSC-95-2411-H-004-028 to I-P Wan.

...I focused on pointing out the fundamental implausibility of generative grammar. By doing so, I hope to divert the attention of as many future linguistic researchers as possible to the more useful and realistic approaches to language, including sociolinguistics, psycholinguistics, applied linguistics, discourse analysis, lexical semantics, functional-systematic (*sic*) grammar, computational linguistics, and so on, and of course, corpus linguistics. (Shei 2004:14)

Shei's advice thus seems to be this: do anything, *except* generative grammar, which is not useful and not realistic. Given that the locally published *Concentric: Studies in Linguistics* is one of the most accessible linguistic journals in Taiwan and that generative grammar is part of virtually every undergraduate and graduate linguistic program in the country, we consider the possibility unsettling, that our future linguists interpret the lack of a formal reply as indifference, concession, or worse still, consent, on the part of generative linguists. The goal of this paper is thus to serve as an antidote. However, we will not pretend to try to definitively demonstrate the plausibility of the generative approach or to be able to convince all readers. If the accumulated research by the generativists in the last fifty years, whose accomplishments many would consider nothing short of spectacular, cannot succeed, this current article certainly will not. Rather, we will clarify some of the generativist views and the current state of affairs and thus point out the inaccuracies and inconsistencies in Shei's argumentation. We will defend the generative paradigm as a worthy scientific pursuit, though fully acknowledging that many issues remain controversial. The reader is urged to keep an open mind and not to assume that a defense of the generative grammar is automatically an attack on functionalism or corpus linguistics. In fact, we advocate the view that the two approaches can and should co-exist.

As a reply to Shei (2004), this paper is organized accordingly. Section 2, responding to Shei's Sections 1 and 2, discusses the shared characteristics in different conceptions of grammar and the nature of grammar as envisioned in generative linguistics. Section 3 examines the evidence offered by Shei from psycholinguistic studies and expands the discussion to psycholinguistic evidence in support of the generative views. Section 4 does the same based on relevant biological studies. Section 5 discusses corpus linguistics in general and also demonstrates its compatibility with the generative approach. Section 6 re-examines the specific corpus data of *huoshi* 火勢 'fire, fire momentum' that Shei used as a case study to demonstrate the concept of 'extended lexical unit'. We will present a case study of our own and argue that corpus linguistics need not be in conflict with generative investigations. Section 7 examines Shei's concluding remarks and also consists of our

conclusion.

2. Grammar and generative grammar

Shei was certainly correct that different people working in different technical domains or different theoretical frameworks might have different conceptions of the term ‘grammar’; however, the term invariably refers to a system of abstract rules or constraints. The American structuralist view is that these are secondary facts to be solely derived from the taxonomic classification of a collection of primary linguistic facts. However, the generative revolt, led by Noam Chomsky, instead championed a rationalist conception of the nature of language system in the Saussurean dichotomy of system versus use. While the latter view has been the mainstream for past several decades, the former view is seeing a revival of a sort, in the study of corpus linguistics especially.

Even though Shei made it amply clear that he is adamantly against ‘generative grammar’, it is not clear at all what precisely he meant by this term. The term ‘generative grammar’ has two technical senses, a narrow sense that refers to a type of (formal) grammar that is ‘generative’, and a broad sense that refers to a linguistic research paradigm that comes with a host of theoretical constructs and hypotheses about human language. The broad, extended sense of the term encompasses the narrow sense, but not vice versa. In this paper, we use the capitalized ‘Generative Grammar’ or ‘generative linguistics’ to indicate the broad sense. The term ‘generativist’ also refers to the broad sense.¹ In 2.1, we will discuss the core technical sense of ‘generative grammar’ and then expand to the several important theoretical tenets of the general generative approach. In 2.2 we examine the innateness issue, and the concept of autonomous syntax is reviewed in 2.3. We will defend the generativist focus on grammatical competence and the pursuit of the universal grammar, in 2.4 and 2.5 respectively.

2.1 The core sense of ‘generative grammar’

The narrow sense is the core of Generative Grammar, but it was overlooked by Shei. In its core sense, a generative grammar is a system of formal and explicit rules that generates all and only the possible sentences in a language. The term ‘language’

¹ In order not to further complicate the issues, we will avoid the term *Chomskyan*. While Chomsky indeed has been the central figure in the development of Generative Grammar, his views are not necessarily accepted by all generativists, as there exist many different frameworks within the generative approach. Thus, this term also seems to have a narrow sense and a broad sense, the former referring to the particular derivational framework that Chomsky advocates and the latter largely equivalent to the term *generativist*.

is used in its mathematical sense here, meaning a set of strings of symbols.² In the case of a human language, the well-formed, or grammatical, strings may be recursive and are thus infinite. The term *generate* is also used in its mathematical sense, meaning *define*. We will illustrate with a simple example. Assuming a language with only two symbols, x and y , and a grammatical string must begin and end with an x , with any number of y 's in between, as summarized in (1). W is the set of symbols and L the set of infinite well-formed strings.

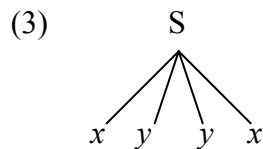
$$(1) \quad W = \{x, y\}$$

$$L = \{S_0, S_1, S_2, \dots, S_n\}, S_0 = xx, S_1 = xyx, S_2 = xyyx, S_3 = xyyyx, S_4 = xyyyyx \dots$$

A generative grammar for (1) may take the form of a single regular expression, as in (2), where the asterisk on y indicates any number of y , including zero.

$$(2) \quad S = xy^*x$$

This grammar is generative in that it generates all and only the well-formed strings in L . However, this grammar, by producing only 'flat' structures, suggests that there is no internal structure in any of the well-formed strings. S_2 is given as an example.

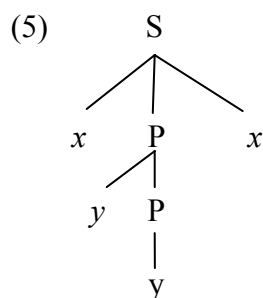


The alternative grammar in (4), which is equally generative in respect to L , on the other hand, indicates a hierarchical structure among the symbols in a string; also, recursion is captured more elegantly by embedding the constituent P within itself. The same string S_2 is given as an example in (5), as a contrast to (3) generated by (2).

$$(4) \quad S \rightarrow x P x$$

$$P \rightarrow y (P)$$

² For example, in his 1984 Nobel lecture entitled 'The Generative Grammar of the Immune System', Niels K. Jerne, Nobel laureate in Medicine and Physiology, equated various features of protein structures with components of a generative grammar.



Thus, both grammars in (2) and (4) are generative and also *observationally adequate*, meaning that they generate all and only grammatical strings in L. However, if the hierarchies reflected in the grammar of (4), as exemplified in (5), indeed better reflect the true state of affairs than (2), then (4) is said to be also *descriptively adequate*, while (2) is not. In the case of a human language, a descriptively adequate grammar is one that goes beyond accounting for just the grammaticality of the data in question and reflects some aspects of the psychological reality of the native speaker.

The generative concept, and practice, of formal and explicit grammars have made great contributions to mathematical linguistics and computational linguistics. Shei's call to abandon the generative approach implies that whatever alternative approaches one adopts, linguistic generalizations need not or should not be stated formally and explicitly. However, given Shei's affinity to corpus linguistics and computational linguistics, he is surely not against this formal aspect of Generative Grammar, which however is by and large rejected, explicitly or implicitly, by most functionalists. Shei does call for grammar to be based on psycholinguistic evidence, which is precisely what the generativists aspire to in their pursuit of descriptive adequacy. So, already there are two crucial aspects of the Generative Grammar that Shei in fact agrees with, despite his call for its demise.

The highest level of adequacy in a generative grammar is nonetheless that of explanatory power, obtained only if it pertains to all and only human languages and thus in some way explains how a child acquires a language. Generative Grammar as a linguistic research paradigm aims at achieving explanatory adequacy. This goal is based on a number of hypotheses. In his article, Shei (2004:2) proclaims to address the implausibility of the following three generativist theoretical tenets: 'Language is innate. Syntax is autonomous. Competence can be separated from performance'.

2.2 The innateness hypothesis

The generativist view holds that the fundamental human capacity to acquire any natural language is innate while the drastically different form-meaning associations of the individual languages must be learned through adequate and appropriate exposure

to a specific speech community. A child can master any of the world's thousands of languages with ease within five to six years, a remarkable feat on any scale considering the enormity of the task. Moreover, once lexical items are set aside and the superficial differences abstracted away, languages only vary within a surprisingly small number of fixed syntactic patterns. The generativist hypothesis is thus that in the course of the evolution of the human species, a (language-specific) biological faculty has developed, which gives rise to the shared properties in all human languages and also helps rule out the infinitely large number of possible and yet non-human languages.³ These general properties are known as the Universal Grammar (UG), which greatly facilitates the child acquiring any human language. Thus, a more accurate statement would be that *language*, or UG, is innate, but *languages*, or grammars of languages, are learned. Note that crucially Shei agrees that *language* is innate, though with certain reservation:

That language is to some extent innate is less controversial. (Shei 2004:2)

To many, including scientists in disciplines outside of linguistics proper, the debate was over. Arguing for sexual choice as the dominant force in shaping the evolution of the human mind, Miller (2000:344)⁴, an evolutionary psychologist, has this to say regarding the innateness debate:

The language theorist Noam Chomsky and other language 'nativists' fought hard against the social science dogma that all human mental abilities are products of learning. It was a heroic fight, but for our purposes all we need to know is that the nativists won.

But to others, the issue remains controversial. For example, *Linguistic Review* published a recent special double issue focusing on innateness (cf. Ritter 2002). We will first address the innateness debate at two levels: a misguided level and an informed level. On a misguided level, opponents argue that language, like other skills, can be acquired without resorting to innate knowledge but say nothing regarding whether language is uniquely human. To Chomsky, the issue is simple: logically, if language is uniquely human, then it is innate, period:

³ As a Christian, the first author of this paper believes that evolution is one of the ways life forms change in the universe that God created and thus that creationism and evolution are not necessarily incompatible.

⁴ The first author wishes to thank Hsin-I Hsieh for bringing his attention to Miller (2000), which is cited in several places in the paper.

If they believe that there is a difference between my granddaughter, a rabbit and a rock, then they believe that language is innate. So people who are proposing that there is something debatable about the assumption that language is innate are just confused. So deeply confused that there is no way of answering their arguments. There is no doubt that language is an innate faculty. (Chomsky 2000)

Thus, any credible evidence against this view might arise from animal language research, especially research with chimpanzees. Claims have been made that a lab-trained Bonobo chimp named Kanzi has demonstrated the communication skills equivalent to that of human toddlers and thus the separation in kind between human and ape language is a myth (e.g. Savage-Rumbaugh and Lewin 1994 and Savage-Rumbaugh 1999). Nativists are quick to dismiss such claims. Here is Chomsky again, in an interview documented in Johnson (1995):

Humans can fly about 30 feet—that's what they do in the Olympics. Is that flying? The question is totally meaningless. In fact the analogy to flying is misleading because when humans fly 30 feet, the organs they're using are kind of homologous to the ones that chickens and eagles use. Arms and wings, in other words, arise from the same branch of the evolutionary tree. Whatever the chimps are doing is not even homologous as far as we know.

This sentiment is not unique to generative linguists. For example, Miller (2000:343), the evolutionary psychologist, expresses the same view:

The ape language controversy was unenlightening because we already knew that chimpanzees do not naturally talk. The fact that they do not suggests that the last common ancestors we shared with chimpanzees, five million years ago, did not talk either. ...there is no more reason to look for rudiments of language in chimps than in baboons, beavers, or birds. The trained use of visual symbols by very clever individual apes like the famous Kanzi is marginal to understanding the evolution of language.

However, here we think generativists, while pursuing their own agenda based on the innateness hypothesis, should give the animal researchers some benefit of the doubt and allow them to pursue a competing hypothesis and see where that leads. Such competition can only be good in that, whether the competing hypothesis ultimately pans out or not, animal language research adds to our understanding of not only animal communications, but also human languages. Note, however, even if

certain animals indeed also possess language to some degree, logically it still does not necessarily negate the innateness hypothesis. This is where the more informed debate on the innateness hypothesis can be found. Since Shei accepts the innateness hypothesis, we shall have no further comments on it *per se* and will move on to the next logical question:

Bates (2003) was not against the idea of language being innate. What ought to be questioned is the form of innateness. ...A distributional view is more likely than a modular one. (Shei 2004:2)

This is a completely reasonable question and is where the real debate should be focused, one that the generativists accept to engage in. Chomsky, for example, fully acknowledges the empirical nature of this debate:⁵

Now a question that could be asked is whether whatever is innate about language is specific to the language faculty or whether it is just some combination of the other aspects of the mind. *That is an empirical question and there is no reason to be dogmatic about it; you look and you see.* What we seem to find is that it is specific. There are properties of the language faculty, which are not found elsewhere, not only in the human mind, but in other biological organisms as far as we know. (Chomsky 2000, emphasis added)

We will defer our discussion on some of the evidence for and against the hypothesis that the language instinct is language-specific and return to this debate in Sections 3 and 4. For now, note that the thesis that language is innate is not a generativist innovation and has in fact been around for a few hundred years. A key source for Chomsky's conception is Rene Descartes, the seventeenth century French philosopher (e.g. Chomsky 1965). More importantly, whether the innate properties of language turn out to be 'specific to a language faculty' or attributable to 'some combination of the other aspects of the mind', Generative Grammar remains intact. Under either the 'modular' view, which Chomsky supports, or the 'distributional' view, which Bates (2003) advocates, it is the innate properties that give rise to the shared properties among languages; thus, the UG hypothesis remains and the pursuit for explanatory adequacy remains. It is illogical to dismiss the generative approach based on the belief that the innate properties of language are not, wholly or partially, attributable to a language-specific biological faculty.

⁵ Pinker's (1994) bestseller offers a host of arguments for the generativist position, but Sampson's (2005) rebuttal has likewise proven popular among non-generativists.

2.3 The autonomous syntax hypothesis

Shei's second statement concerns the autonomous syntax hypothesis. It has indeed been recognized that one of the characteristics that unite the various grammatical frameworks, e.g. Transformational Grammar (TG) and its later incarnations such as the Government and Binding Theory (GB) and the Minimalist Program (MP), Relational Grammar (RG), Head-driven Phrase Structure Grammar (HPSG), and Lexical-Functional Grammar (LFG), to name just a few, under the Generative Grammar banner is the assumption that syntax is largely autonomous in that syntactic generalizations form a self-contained system, parallel to phonology for example (e.g. Newmeyer 1991).

As will become clear, Shei's call to abandon the generative approach boils down to his utter rejection of autonomous syntax and his enthusiasm for corpus linguistics, which, in Shei's view, is subsumed by functional linguistics. Thus, we should quickly point out that the hypothesis that syntax is a self-contained system of *internal* constraints, logically, does not at all negate the thesis that such syntax-internal constraints find syntax-external motivations. It is true that most generativists appear to be indifferent to this fact, but others have fully accepted it. Here we quote two of the most prominent generativists:

Surely there are significant connections between structure and function: this is not and has never been in doubt. (Chomsky 1975:56)

...I regard the assumption that much of grammatical structure is motivated by external functional pressure as being a fairly uncontroversial one, even among the most doctrinaire formal linguists. (Newmeyer 2003:687)

Thus, whether syntax is autonomous should be judged, primarily, if not solely, on the basis of whether syntactic principles are self-contained and systematic. That some, or perhaps even many, of these principles may find system-external motivations is beside the point.⁶

The functionalist anti-autonomous syntax view should apply equally to phonology, which operates under a precisely parallel assumption that phonological regularities and constraints form a self-contained system, independent of other linguistic modules. Bybee and Hopper (2001:3), for example, contend that "grammar" itself and associated theoretical postulates like "syntax" and "phonology"

⁶ Newmeyer (1991, 2003) provides examples such as the chess game and bodily organs to illustrate the point we are making regarding autonomous syntax.

have no autonomous existence beyond local storage and processing'. Curiously, in functionalist literature the same level of objection to autonomous syntax is not found towards autonomous phonology. This indicates that it is not the concept of autonomy or modularity that many functionalists disapprove of; rather, it is autonomous *syntax* specifically. Furthermore, judging from their frequent criticisms of grammar as an abstraction removed from meaning, it is evident that this double standard is due to the perception that syntax, compared with phonology, has a closer affinity to semantics. Here we will make two points. One is about the separation of syntax and semantics in structuralism in general, and the second point relates to the place of semantics in Generative Grammar.

The isolation of syntax from semantics, or the concept of autonomous syntax, is not a generativist innovation and in fact finds its roots in structuralism. The underpinning assumption of structuralism in linguistics is that underlying the observable phenomena of language there are discrete, determinate entities analyzable as systems of correlations between forms and meanings. Thus, the inheritance of this autonomous thesis puts Generative Grammar squarely within the structuralist vein of linguistics. However, Shei, like many other critics of generative autonomous syntax, shows a conflicting sympathy to structuralism:

It is the generative grammarians who drove Bloomfield and the structuralists (who should in fact be credited for their comprehensive fieldwork) away...and confined themselves in the ivory tower of syntax to whom the label 'narrow scope' seems better suited. (Shei 2004:14)

Note that Shei's alliance is with the American structuralists no less.⁷ But one thing that separates Saussurean structuralism from the Bloomfieldian structuralism is the ontological status ascribed to the syntactic system. While the former holds it to be psychologically real in the collective mind of a linguistic community, the latter analyzes linguistic forms without consideration as to whether the analysis may be psychologically real. Bloomfield himself was a professed behaviorist (e.g. Crystal 1997:408). Given Shei's frequent reference to psycholinguistic evidence in his argumentation, the psychologistic view of autonomous syntax and the Saussurean distinction between *langue* and *parole*, which the Generative Grammar revived, should at least have been the lesser of the two evils for him. Thus, clearly, Shei's

⁷ Shei and other supporters of Bloomfieldian structuralism may be dismayed to learn that Bloomfield published an article in 1939 outlining a generative grammar similar to the more detailed one Chomsky put forth in his 1947 BA thesis. Bloomfield's article, entitled, *Menomini Morphophonemics*, was published in the *Czech Travaux du Cercle Linguistique de Prague* in 1939 (Barsky 1997, Chp 4, Sec. 4).

affinity to the (Bloomfieldian) structuralists rests solely on their inductive and taxonomic methodology based on items in a corpus, which the generativists oppose. We will address this particular issue in Sections 5 and 6.

Our second point is that the void of meaning in generativist grammars has often been exaggerated and has become more a stereotype. It is true that early transformational models indeed followed the structuralist ways in factoring out meaning in syntactic theorizing. Later models, however, do, to a varying extent, take into account semantic notions. The short-lived Generative Semantics, for example, also enjoyed a resurrection of a sort in Baker's (1988) theory of incorporation, which has been well-received within the derivational framework. The theta-theory, in relation to the well-known problem of linking semantic roles in event-structures to syntactic arguments, is another good example. In the Minimalist pursuit, what is traditionally considered semantics is taken to be the part of syntax that is close to the interface system that involves the use of language (Chomsky 2000). Some non-transformational generative frameworks also incorporate semantics with syntactic representation, e.g. HPSG, Montague Grammar, and Categorical Grammar.

And then there are notions of pragmatics and discourse. The stereotypical characterization by the functionalists is again that they have no place whatsoever within the generativist approach. That is not entirely true either. While generative frameworks in general factor out discourse and pragmatic considerations in their grammatical descriptions, notions such as 'topic' and 'focus', for example, have been incorporated in the derivational framework, as the functional categories of Top and Foc, as well as in the non-derivational LFG, as the grammaticalized functions TOPIC and FOCUS. To the best of our knowledge, no generativist has ever made the claim that no syntactic rules, constraints, or principles, whether specific to a language or applicable to UG, came to the present shape, partially or entirely, due to external forces. The reason no such claim was ever made is simple: external motivations do not threaten the Generative Grammar or the autonomous syntax hypothesis. Likewise, many functionalists readily acknowledge that there are syntactic processes that simply cannot be attributed to syntax-external forces. Such syntactic processes are said to be 'arbitrary' and grammaticalized. Thus, grammaticalization seems to be a common ground between the generativists and the functionalists. A question for the functionalists to ask themselves is this: is it not a worthwhile research project to find out whether or to what extent the individual arbitrary or grammaticalized syntactic phenomena, together with the ones that have been recognized as functionally motivated, form an internally coherent system? If one answers positively, then one must accept autonomous syntax as a reasonable working hypothesis, besides that of functionalism.

2.4 The competence hypothesis

The generativist goal is indeed not to describe linguistic performance, or the actual speech utterances that a speaker produces or the actual linguistic data collected in a corpus of a particular language community. Rather, Generative Grammar attempts to describe linguistic competence, the *tacit* grammatical knowledge that a native speaker possesses which enables her to produce grammatical sentences and distinguish them from ungrammatical ones. Performance, however, is the empirical and formal realization of competence. The generativist view is thus that grammatical competence can, and in fact should, be studied independently of considerations of performance, which is more of concern in other disciplines such as pragmatics, discourse analysis, literary theory, physiology, psychology, and neurology. No generativist that we know of has ever denied the importance of any of these disciplines.

Thus, while the focus of generative syntactic pursuit is squarely on grammatical competence and competence only, competence and performance are never disconnected in the generativist view. People's interests vary. It is one thing for one to criticize the group of linguists of poor taste for paying little attention to what one might consider the more interesting issues regarding language use, performance factors, or functional motivations, it is quite another to deny the plausibility of linguistic competence and thus the legitimacy of generative syntax. The two are unfortunately often confused. Logically, the competence hypothesis is entailed by the innateness hypothesis, as competence in this sense is not something learned; rather it is innate and triggered by the exposure to a language. Given Shei's endorsement of the innateness hypothesis, his rejection of the competence hypothesis can at best be interpreted as an accusation against the generativists of poor taste.

2.5 The UG hypothesis

Also entailed by the innateness hypothesis is the existence of UG. Whether biologically the innate properties are specific to a language faculty or are entirely attributable to a combination of other cognitive factors or are something in between, these innate properties universally constrain child language acquisition and give rise to the universal characteristics in human languages. Thus, they are independently worthy of study as such. Again, it is one thing to accuse the group of linguists who pursue UG based on data from primarily one language, e.g. English, it is quite another to dismiss UG itself. Again, it is inconsistent for Shei to accept the innateness hypothesis and yet deny the legitimacy of the generativist pursuit of the universal properties of whatever is innate about language:

Although universal grammar is the basic rationale of Chomskyan syntactic theories, there is no conclusive support for these theories in languages other than English, where Chomskyan theories originated. (Shei 2004:14)

In this statement quoted above, Shei contradicts himself in giving too much and too little credit to generative theories at the same time. If Shei truly believes that evidence from English supports a UG-based theory, then surely we should try to extend it to other languages. However, no fair-minded generativist would make such a strong claim that there is *conclusive* support, in English or any other language, for any of the UG-based theories. Chomsky, for example, has stated quite plainly that the Minimalist Program is just that, a research program, even though it has had its successes:

This *Principles and Parameters* approach may or may not turn out to be justified; one can never know. But as a research program, it has been highly successful, yielding an explosion of empirical inquiry into a very wide range of typologically varied languages... (Chomsky 2006:xii)

Thus, Shei gives too little credit to generative theories by perpetrating the myth that generative findings are based on English only. It is well-known that Chomsky has quipped that anything you find in one language can also be found in every other language, at a more abstract level of representation (Pinker and Bloom 1990). Thus, in theory, insights to UG may indeed come from the study of a single language; however, that is never the case in practice. Commenting on Givón's (2002) accusation that the generativists failed to pay serious attention to diversity⁸, Everett (2005:163), whose overall sympathy is with the functionalist, concedes frankly that 'there is too much high-quality and empirically rigorous work in generative grammar, not to mention many unempirical functional works, to apply it globally'. Generative grammarians have indeed studied hundreds of languages and put various hypotheses to test, whether these hypotheses arise due to theoretical or formal considerations⁹ or

⁸ As the burden of proof rests with the accusers, such claims should really be backed up by more concrete evidence or at least some kind of statistics. We are not aware of any such studies.

⁹ As pointed out in Her (2005), the constant drive for economy and simplicity has been a significant motivation for the successive evolution of the earliest Transformation Grammar to the current Minimalist Program. The general X-bar scheme that replaced the stipulated phrase structure rules and the single operation of *Move- α* generalized from the various construction-specific transformations are two good examples. In certain versions of the Minimalist Program, these two are further reduced to *Merge* and *Agree*, the only syntactic operations.

empirically derived from linguistic facts in a certain language¹⁰. It is thus a fallacy that generativists dream up UG or universal principles without first observing languages:

Unlike generative grammarians who use more of a deductive approach (they, for example, *contrived UG first and then went on to examine whether sentences follow UG principles*), corpus linguists use an inductive approach and start from observations of language in use. Thus the observed rules in corpus linguistics tend to be locally relevant and less pompous looking. (Shei 2004:9, emphasis added)

First of all, for argument's sake, let's say that Shei and others like him are right and UG is contrived. As long as languages do follow universal principles, the concept of UG is validated. It matters little whether the concept itself is contrived first and then validated or universal properties are observed first before UG is proposed. But the fact is that the concept of UG is far from being contrived. Rather, it is a consequence of the innateness hypothesis, which is in turn entailed by the scientific and thus falsifiable observation that *language* is uniquely human. It is further supported by a host of observations on language acquisition, such as lack of overt instruction, poverty of stimulus, uniformity in language acquisition, creolization, and the presence of language universals. Furthermore, language universals or universal principles all came about from observations of natural languages and are thus fully falsifiable. Biological and psycholinguistic evidence will be discussed in the next two sections.

An element in the generative movement that its critics consistently overlook is the element of competition from within. Universally relevant, or 'pompous' as Shei calls it, claims, especially when made by influential figures, such as Chomsky himself, are regularly challenged, often by the claimant's own disciples, if not by proponents of competing frameworks within the generative paradigm. Such internal scrutiny is far more stringent than that of by opponents of UG, and most certainly taken far more seriously. The continuous evolution of the generative paradigm and the diversity of views within are indicative of this vigorous and rigorous scientific process.

Before moving on to examine the relevant psycholinguistic evidence, we shall have our final remark on UG in relation to culture:

Although the innate hypothesis can be supported, grammar is constantly shaped by culture and interpersonal interactions. (Shei 2004:1)

¹⁰ A good example is James Huang's (1982) study of Chinese that led to the analysis of WH-move in LF and thus the parameter of overt versus covert movement.

Here Shei confuses the innate UG, which he confesses to supporting, and the individual grammars. This widespread confusion is precisely what Derek Bickerton, a leading researcher on language evolution, has cautioned in a recent article:

Of course it (language evolution) has stopped, because the biological evolution of humans... has, to all intents and purposes, stopped also. What is happening (and has been happening for perhaps as many as a hundred thousand years) is cultural change...; within the envelope of the language faculty, languages are recycling the limited alternatives that this biological envelope makes available. It should always be a warning signal when writers engage in the kind of sleight-of-hand that persistently switches between 'language' and 'languages'. (Bickerton 2007:511)

Thus, it would have been more accurate, had Shei stated that *grammars* (of individual *languages*) are constantly shaped by culture. UG itself, on the other hand, is completely insulated from culture and personal interactions. All changes in grammars, whatever the cause or motivation, happen only within the parameters allowed by the innate UG. That an individual grammar can be shaped by communication has never been in doubt. Chomsky's own comments leave no room for doubt:

Searle argues that 'it is reasonable to suppose that the needs of communication influenced structure'. I agree. (Chomsky 1975:58)

3. Psycholinguistic evidence

In the classical model of Wernicke-Lichtheim, specific locations in the brain are associated with specific linguistic functions or language modules. This model thus indicates a close connection between brain lesions and different types of aphasia. However, Shei cited Dingwall's (Dingwall 1998:92-94) study indicating not only that Broca's aphasia, where sentential structure is often ill-formed as in the telegraphic speech with meaning fairly intact, is not necessarily related to injury to Broca's area, but also that lesions in Broca's area do not necessarily produce Broca's aphasia. Shei then went on to discredit the existence of specific language impairment (SLI), citing Bates (2003), who instead attributes SLI to general deficit in auditory processing, and Cowley (2001), who claims that SLI sufferers are in fact mentally subnormal. Based on this, Shei claims that psycholinguistic evidence puts the concept of autonomous syntax in question and as a consequence generative linguistics is in doubt, because its 'research methodology depends on it' (Shei 2004:4). We will first discuss the

psycholinguistic evidence from studies on aphasia and SLI before examining the logic in Shei's argumentation here.

3.1 Aphasia, SLI, and the language module

Available psycholinguistic evidence in studies of either aphasia or SLIs is not as one-sided or simplistic as Shei portrays. More recent studies have found that syntactic processing is indeed involved in Broca's area (e.g. Grodzinsky 2000, Moro et al. 2001, Müller et al. 2003)¹¹. Thus, the indications are that under normal brain development, each of the cell areas is involved in performing specific functions and thus a close relationship between cortical regions and language modules or functions can be justified. More specifically, an aphasic syndrome can be made up of a cluster of symptoms, each of which may have a different locus of damage. For example, apraxia, an impairment in the sequencing of speech sounds is caused by damage to the pre-frontal gyrus of the insula, and morphosyntactic comprehension deficits may be caused by damage to the anterior portion of the temporal lobe. Precise localization for deficits in morphosyntactic production has yet to be determined, but the arcuate fasciculus seems to be involved. It is not clear what functions can be attributed to Broca's area *per se*. Evidence from the performance of brain-damaged patients convinces Ullman et al. (1997) that dominant hemisphere anterior neocortical areas (Broca's area, area 44, 45)¹² subserve the computation of regular grammatical forms, while dominant hemisphere posterior neocortical areas (area 39, 40) subserve the computation of irregular grammatical forms; in addition, they argue that the basal ganglia contribute to the computation of regulars but not irregulars.¹³

A fair statement is thus that the localization issue still remains largely unresolved, in spite of the numerous studies examining data from grammatical analysis, computing modeling, on-line language processing, language acquisition, and language disorders. It is therefore not yet clear whether syntax, which the generativists claim to be an autonomous rule system, is associated with an equally autonomous and identifiable area in the brain. Nonetheless, a basic tenet of linguistic aphasiology is that the mind is composed of a set of processing elements, modules, which are dissociable, a view compatible with generative linguistics.

¹¹ We thank the anonymous reviewer for the Grodzinsky (2000) reference.

¹² The area numbers refer to the cytoarchitectural maps of the human brain developed by Brodmann (1909) who found that there are 46 different kinds of neurons in the cortex, and each of them is located in groups.

¹³ As an anonymous reviewer correctly points out, this work does not support localization as it argues that basal ganglia are also involved in syntax.

Regarding genetically inheritable linguistic impairment¹⁴ or SLI, skeptics, such as those cited in Shei (2004), insist that it is usually accompanied with low performance in some other areas, such as motor skills, auditory processing, or intelligence and thus attribute it to either non-linguistic factors or a more general deficit in mental capacity. In other words, they fundamentally reject the existence of SLI. However, as the name ‘specific language impairment’ dictates, its diagnosis is necessarily based on exclusion: a delayed language development with no showing of hearing loss, mental retardation, or emotional disorders. Cheung (2003), for example, in a study of Mandarin-speaking SLI children in Taiwan, concludes that there is no significant difference in their memory for verbal stimuli. Simpson and Rice (2004) stress the crucial fact that a child with SLI does not have a low IQ or poor hearing and that SLI, as a language disorder, can be diagnosed precisely and accurately. A good example is the Rice-Wexler Test of Early Grammatical Impairment, designed for children ages 3 to 8. If Shei and the skeptics are right, then every single case of SLI that was ever diagnosed, investigated, and documented in the last forty years by numerous speech pathologists and neurolinguists would have been a mistake. Evidence from SLI clearly leans towards the disassociation of language and intelligence (e.g. Rice 2002), contrary to the picture of controversy Shei paints.¹⁵

3.2 Arguments in favor of the language module

If the generativist hypothesis is correct, that language and intelligence are largely dissociable, as SLI indicates, then logically there should be pathological cases where prolonged delay in mental development is unaccompanied by significant deficit in language capacity. Indeed there are: the Williams syndrome. This, Shei fails to mention.

To demonstrate the disassociation of general intelligence and language, we cite only a more drastic case in the *savant* Christopher, reported in detail by Smith and Tsimpli (1995). This brief description by Smith (1999: 24) will suffice:

Christopher is a flawed genius. He lives in a sheltered accommodation because he is unable to look after himself. He cannot reliably find his way around, and he has poor hand-eye coordination, so that everyday tasks like shaving or doing up his buttons are tedious chores. He fails to conserve number (a task which is within the capabilities of most five-year-olds), and he has many of the characteristics of

¹⁴ It is fairly uncontroversial that SLI is likely attributable to genetics (e.g. Bickerton 2007). Even a harsh critic to the generative framework like Sampson (2005:123) agrees.

¹⁵ The reader is referred to Jenkins (2000) for a good review on SLI. Again, we thank the anonymous reviewer for providing the Jenkins (2000) reference.

autism. Yet he can read, write, translate, and communicate in some fifteen or twenty languages.

Given the demonstrated disassociation of intelligence from linguistic ability in such cases and Williams syndrome children, the existence of SLI can only be expected. Likewise, it can only be expected that delay in development be found in both intelligence and linguistic ability, such as in the Down's syndrome. Furthermore, the fact that language ability can be damaged largely in isolation from other mental or cognitive capabilities, as in cases of aphasia, would suggest that language ability alone is subject to genetic abnormality, thus SLI. Clearly then, even though a definitive answer is still premature, there is ample psycholinguistic evidence supporting the generative paradigm as a credible and worthy scientific pursuit.

More importantly, one should remember the fact that the primary goal of generative linguistics is to provide a model where a constrained set of possible rules can be formulated formally and accurately to account for the boundless creativity of human language. Such rules must be learnable; that is, they must be formulated in a way that accounts for how a child is able to master any language in such a short period of time with little or no explicit training. They must also fulfill the goal of universality: they must capture common features of the grammars of all languages. Such a formal model and explicit formulations make testable claims about whether certain sentences might be more difficult than others for speakers to produce or understand.¹⁶ It also makes predictions about how children might move through stages in learning languages, some of which were substantiated by studies of children's language development, e.g. Ratner, Gleason, and Narasimhan (1998). Thus, whether syntax forms an autonomous system depends on whether syntactic rules by and large form a unified and self-contained system *internally*.

An internally congruent and autonomous abstract system may indeed be neurologically located in a centralized region, but it may also be the concerted function of brain cells in distributed areas, neither of which is a logical necessity. Therefore, whether the classical Wernicke-Lichtheim's model of brain areas is correct or not, the clear distinction between Broca's aphasia, where syntax suffers, and Wernicke's aphasia, where semantics is impaired, is a significant indication of the disassociation between grammar and meaning. Shei fails to recognize this in his argumentation.

¹⁶ For example, psycholinguistic studies on empty categories by Bever and McElree (1988) and McElree and Bever (1989) are such endeavors to empirically test generative claims. Again, we thank the anonymous reviewer for providing these two references.

4. Biological evidence

At the biological level, it should be no surprise then that UG, as a self-contained mental system, is not due to a single gene or even an exclusive group of genes. Even though the Vargha-Khadem team has identified a mutation of FOXP2 exclusively in the SLI members in the KE family in the UK, the language impairment is accompanied with lower intelligence, compared with non-SLI family members (Sampson 2005:124), and, furthermore, it remains unknown what other genes FOXP2 turns on or off and what functions they serve. Thus, the popular media's characterization of the FOXP2 gene as 'a language gene', or worse still '*the* language gene', is much too premature, and quite likely just wrong (e.g. Bickerton 2007, Miller 2000:23, and Sampson 2005:124). Yet, according to the survey by Cavalli-Sforza et al. (1988), there is indeed a close correspondence between the genetic tree of human races and the tree of language families, hinting at least indirectly a link between genes and language.

It is thus fair to say that genetics so far provides no direct evidence for UG. However, in his Section 4, which is entitled 'biological basis', Shei does not dispute the innateness hypothesis, nor does he deny the existence of a biologically endowed UG. Rather, he cites Ji (1997) and Snow (1996) to demonstrate the importance of environment and culture in the development of language. Before we look at what specifically Ji (1997) and Snow (1996) argue for, we shall first repeat the crucial point made towards the end of Section 2.4, that the generative linguists fully acknowledge the influence of culture in shaping linguistic structures, but UG itself is insulated from such factors. Studies on children's language acquisition have demonstrated that language acquisition, cross-linguistically, follows a highly similar order in terms of the acquisition of various syntactic constructions. If social, cultural, and other environmental factors do affect UG, we would expect to find as many different variant orders as there are different registers and different cultures. This is not the case.

Note that Ji (1997) in fact claims that cell language, where molecules can be viewed as signs and the messages conveyed are gene-directed cell processes analogous to replication and translation in language. He thus proposes that the natural component of human language is encoded in DNA and that there is an isomorphism between cell language and human language. What is particularly 'wise' in Ji's biological model of language, according to Shei, is the inclusion of the influence of culture in this model. However, note that Ji does *not* suggest that culture has anything to do with UG directly. Instead, he proposes that an appropriate culture and environment is necessary for the language-enabling brain structure to give rise to UG.

Nothing in this proposal is incompatible with generative linguistics, which hypothesizes that all healthy human newborns are endowed with UG and will develop language under normal circumstances. It is thus a possibility within generative linguistics that the failure of a wild child adopted at birth by apes to develop language is due partly to an underdeveloped UG and partly to lack of language exposure. Ji's observation that human language consists of two components: cultural and natural, which again is characterized as 'wise' by Shei, is simply common sense.

Likewise, Snow's (1996) view, which Shei cites to discredit generative linguistics, is in fact not incompatible with generative linguistics. Few would disagree with her view that a well-designed brain and a well-designed environment are both necessary for normal language development. However, she did not take the cultural and environmental factors and argue against innatism. Her caution about the confusion of biology with genetic heredity again seems like common sense, especially when biology is said to include pathology, as Shei (2004:7) argues.

Let's see what consensus evolution scientists have reached. Studies in evolutionary psychology have demonstrated that humans have far more instincts than other animals and the human brain has many more special devices for learning, as a result of evolution. Besides language, other seemingly 'cultural' aspects of human capacity such as morality, humor, art, and music can also be attributable to adaptation, due to either survival benefits or sexual choices or both (e.g. Miller 2000). Therefore, if language is indeed uniquely human, as it surely appears to be, the hypothesis that there is a special language acquisition module is reasonably worth pursuing. It has specific features and capacities, and most likely a definite neural organization. Miller (2000:345) sums it up this way:

The current debate no longer concerns whether language is an adaptation, but what it is an adaptation for.

Chomsky, though not accepting the adaptationist position *per se*, proposes a more encompassing view. He concludes that there are factors in three areas that affect the growth of language in an individual, given that the language faculty must share the general characteristics of other biological systems (Chomsky 2006:180):

- (1) Genetic factors, apparently near uniform for the species, the topic of UG. The genetic endowment interprets part of the environment as linguistic experience, a non-trivial task that the infant carries out reflexively, and determines the general course of the development the language faculty to the languages attained.

- (2) Experience, which leads to variation, within a fairly narrow range, as in the case of other subsystems of the human capacity and the organism generally.
- (3) Principles not specific to the faculty of language.

Clearly, none of the views expressed by Miller, Snow, or Ji, contradicts the three kinds of factors stated by Chomsky. Once again, the point we wish to make here is a humble one: evidence from biology, genetics, and evolution science does not show that Chomsky and the generativists are on a track which is so wrong that it ought to be abandoned. If Shei's intention is purely to demonstrate the importance of culture and environment in language development, not to bring the generative framework to its demise, we will have no disagreement.

5. Corpus linguistics versus generative linguistics

Shei's central mission is to convince future linguists not to explore the generative approach and to adopt a corpus-based approach instead. But, is the generativist pursuit of grammar fundamentally in conflict with the investigations into how a language is actually used, by an individual or in a speech community, as Shei contends?

Corpus linguistics is the opposite of generative grammar... Unfortunately, the two approaches do not converge and compensate for each other. (Shei 2004:8)

This is not a view shared by all, or even most, corpus linguists. For example, Biber et al. (1998:10, 271) insist that corpus-based analysis should not be taken as the single correct approach and that the Chomskyan approach and the corpus approach can be seen as 'complementary rather than conflicting'. The sort of negative sentiment shared by Shei is, perhaps in part, due to the (mis)perception that generative linguists in general rebuff the entire corpus-linguistic enterprise. This is what Chomsky once said in an interview (Aarts 2000:5):

Bas Aarts: What is your view of modern corpus linguistics?

Noam Chomsky: It doesn't exist.

Such rhetoric, from either side, is certainly not conducive to efforts at reconciliation. Let's examine the central difference between the two approaches. Shei is quite correct that the two approaches differ in methodologies: the generative linguist works primarily with introspective data intended to reflect an idealized state of grammatical competence and the corpus linguist focuses on actual performance of

data in language use. More importantly, the two approaches also differ in their objectives. While the former aims to reveal the *knowledge of language* in the mind, the latter seeks to describe how *languages* are actually used. Biber et al. (1998:1) divide linguistics into two main persuasions: generative linguists, who look at what is *theoretically possible* and descriptive linguists, who study the naturally occurring, actual language use. Corpus linguistics is thus more interested in what is *practically probable*, a view shared by Kennedy (1998:270-273). In Chomsky's more recent work (e.g. Chomsky 1986), language exists in two forms: I-language and E-language. I-language is the knowledge of language, internally represented in the brain of an individual; E-language exists in the external utterances in the arena of use. It is a fair statement that generative linguistics focuses on I-language; corpus linguistics, E-language. We will take it for granted that UG is not going to emerge from corpus studies alone and generative studies will not reveal any significant aspects of language use *per se*.¹⁷ Therefore, a rather crucial question remains: do the differences in objectives and methodologies make the two approaches entirely incompatible? We can approach the question from two opposing perspectives: 1) can generativists benefit from language corpora and corpus-based studies? And 2) Can corpus linguists benefit from introspective data and generative findings?

To the first question, some generativists' answer is decisively positive and we agree. Fillmore (1992:35), for example, acknowledged that corpora allow the establishment of new facts, some of which one 'couldn't imagine finding out about in any other way'. In fact, the disparity between introspective data and collected corpora may turn out to be not as great as commonly assumed. After all, introspective data must involve grammaticality judgment, performed by native speakers, and are thus performance data (Schutze 1996). Logically then, nothing should prevent a generative linguist from using other types of performance data, however collected, as long as native intuitions confirm their (un)grammaticality and the working linguists can see the insight they reveal. Thus, no generativist would disagree with the following statement by Neil Smith, a devoted supporter of Chomsky, regarding generative research:

While idealization is necessary, it must be emphasized that idealization away from speech errors, for instance, still allows one to use performance mistakes such as slips of the tongue as *evidence*... All our understanding of linguistic knowledge...has to be supported by evidence, and where that evidence comes

¹⁷ Now we see that Chomsky was being honest and straightforward, rather than intentionally provocative, because, to him, linguistics should be about the pursuit of I-language (e.g. Chomsky 1986). But Chomsky, or any other single individual, does not have a monopoly on defining linguistics. And the reality of the field certainly does not reflect this exclusivist view.

from is limited only by our imagination and ingenuity. (Smith 1999:15, emphasis in original)

This view is likewise shared by many advocates of corpus linguistics. Kennedy, author of *An Introduction to Corpus Linguistics*, states:

The use of corpus as a source of evidence however is not necessarily incompatible with any linguistic theory, and progress in the language sciences as a whole is likely to benefit from a judicious use of evidence from various sources: texts, *introspection, elicitation or any other types of experimentation as appropriate.* (Kennedy 1998:8, emphasis added)

By the same token, corpus linguists should not reject introspective data, and many of them agree. Biber et al. (1998:10) fully acknowledge that intuition can indeed lead to interesting corpus-based studies and corpus-based research questions in fact often find their origin in theoretical studies. Kenney (1998:271-2), conceding the limitations of corpora, advocates the same view:

The use of both introspection and corpus-based analysis can contribute to linguistic analysis and description. Corpora cannot tell us everything how a language works. For example, they cannot be used as a basis for stating what structure or processes are not possible... The fact that an item or structure does not appear in even the largest corpus does not necessarily mean that it cannot occur, but could suggest that the corpus might be inadequate or the item infrequent. (Kennedy 1998:8)

Given that corpora can never fully encompass any language, all of which are infinite in nature, introspective data must always have its role in any kind of linguistic exploration. Clearly then, generative linguistics and corpus linguistics are not as diametrically opposed as Shei tries to have our future linguists believe. In fact, we shall further demonstrate that even in terms of methodology and objective the two kinds of linguistics may also be complementary to each other. This is what Chomsky had to say on a different occasion:

Chomsky once said in a class lecture (I am sure he'd said it many times) that it would be a mistake to come up with a grammar of English full of lots of rules and little riders that got all the facts right, down to every detail. The reason it would be wrong is not that it would not be an honorable scientific endeavor, but rather that

you'd be so bogged down in little details that you'd find nothing of sufficient generality that would lead you to make hypotheses about UG. (Sells 1985:27)

Corpus linguists are indeed not interested in hypotheses about UG and some, like Shei, may even question its very existence. Thus, Shei fully acknowledges that the observed rules in corpus linguistics are in general only locally relevant and quotes Meyer to support this fact:

[A] very common use of corpora: to provide a detailed study of a particular grammatical construction that yields linguistic information on the construction, such as the various forms it has, its overall frequency, the particular contexts in which it occurs... and its communicative potential. (Meyer 2002:12)

The grammars covered in corpus-based investigations are often concrete and specific and by no means intended to be all-encompassing, or “universal.” Little attempt has been made to explore the “core” of grammar of a highly abstract nature. (Shei 2004:7)

If Chomsky does, like any reasonable linguist would, consider such detailed studies ‘an honorable scientific endeavor’, then it is a matter of personal taste in terms of the object of study, not a choice between right and wrong. To Chomsky, and most generativists, the pursuit of principles of UG is certainly far more meaningful, but we doubt very much that most of them would not find the detailed study of *risk* and *let alone* interesting (cf. Fillmore 1992 and Fillmore et al. 1988) or would object to corpus-based study on the subtle differences in the syntax and semantics in near-synonym pairs such as 高興-快樂 *gaoxing-kuai* ‘happy’, 累-疲倦 *lei-pijuan* ‘tired’, and 勸-說服 *quan-shuifu* ‘persuade’ (cf. Tsai 1998 and Tsai et al. 1998). UG certainly cannot emerge from studies of corpora alone, and the pursuit of UG cannot reveal the detailed uses of language-specific constructions and lexical items. However, it is unthinkable that a linguist of any persuasion would consider it unworthy to come up with a grammar with reasonably broad coverage for any language. But common sense should dictate that such a project utilize both known universal principles and detailed studies on important lexical items and constructions in the language. Therefore, the fact is many generative linguists are also interested in formulating concrete detailed rules for language-specific constructions. One might even argue that in the early days of the Transformational Grammar that was precisely what most generative linguists did (e.g. Goldberg 1995:1).

Many critics of the generative movement, like Shei, equate the particular branch of generative syntax that Chomsky advocates with the entire generative enterprise. This mistake, while partly attributable to the awesome influence Chomsky has had over the field, reflects these critics' failure to see the rich diversity within the generative movement. Commenting on Chomsky and his approach to language, Crystal (1997:409) writes:

Since the 1950s, much of the linguistics has been taken up with proposals to develop the form of generative grammars, and the original theory has been reformulated several times. During the same period, also, there have been several alternative models of the grammatical analysis to those expounded by Chomsky and his associates, some of which have attracted considerable support. As a consequence, linguistic theory, the core of scientific language study, is now a lively and controversial field.

Within the tradition of Generalized Phrase Structure Grammar (GPSG), which later developed into Head-driven Phrase Structure Grammar (HPSG), concrete detailed analyses are considered a virtue (e.g. Watson 1985:198). Likewise, Starosta (e.g. 1987, 1988), who considered his Lexicase framework one of the truest generative theories, also shared this attitude (Her 1991:10). Bresnan, one of the chief theorists of the LFG framework, has been advocating stochastic implementations of an Optimality-Theoretic syntactic model based on LFG, in order to account for language-specific and cross-linguistic statistical regularities (cf. Sells 2001, Bresnan and Aissen 2002, and Newmeyer 2003:683). Perhaps the best examples are found in the Construction Grammar, which clearly demonstrates the possibility of harmonious co-existence between a generative framework and corpus linguistics, as works within this framework routinely refer to corpus data (cf. Goldberg 1995, Kay and Fillmore 1999, and Goldberg and Jackendoff 2004, among others).¹⁸ The point is thus again this: the division between generative linguistics and corpus linguistics, in reality, is not as black-and-white as Shei has characterized.

¹⁸ Construction Grammar is generative, but, like LFG, HPSG, Lexicase, and other non-Chomskyan generative theories, it is not transformational. Goldberg (1995:7) states explicitly:

Construction Grammar is generative in the sense that it tries to account for the infinite number of expressions that are allowed by the grammar while attempting to account for the fact that an infinite number of other expressions are ruled out or disallowed. Construction Grammar is not transformational.

Corpus linguistics does not have such a goal. Yet, the studies conducted within Construction Grammar rely heavily on corpus data.

Finally, let's look at the issue of methodology. While generative linguistics insists on the hypothetico-deductive method (e.g. Chomsky 1957 and Starosta 1987), corpus linguists pride themselves on employing the inductive method and focus on observation of language in use leading to theory (e.g. Leech 1992:107). Surely, the difference appears to be diametrical and irreconcilable. That certainly seems to be how Chomsky sees it:

You don't take a corpus, you ask questions. You do exactly what they do in the natural sciences. ... You have to ask probing questions of nature. That's what is called experimentation, and then you may get some answers that mean something. ... You can take as many texts as you like, you can take tape recordings, but you'll never get the answer. (Chomsky, quoted in Aarts (2000:6))

Thus, it does seem that the generative approach is top-down, inside-out, and theory-driven; the corpus approach is bottom-up, outside-in, and data-driven. The two approaches, on the theoretical level, are thus truly competing paradigms. The question is this: is such competition unhealthy to the science of language? Absolutely not. Competing frameworks within the generative paradigm itself, as Crystal (1997:409) has pointed out, contribute to the liveliness of the field.¹⁹ One might argue that in the scientific pursuit of truth competing alternatives are crucial, if not necessary, on every level, be it theory, method, or analysis. Opponents may criticize or attack their rivals, but as long as this is done according to scientific principles, such rivalry is healthy in the interest of truth. Given the liveliness of both generative linguistics and corpus linguistics, the obvious fact is that this healthy rivalry will go on for some time to come before one of them is defeated.

Then, again, the two kinds of linguistics may reconcile. After all, in practice, one is hard put to find any theoretical advance within generative linguistics, including the creation of the theory itself in fact, that did not start out from observations of language data. In other words, whatever 'probing questions' the generativists may ask and whatever experiments, based on introspective data, solicited data, acquisition data, speech error data, or whatever, they conduct, they are always based on hypotheses or insights derived from previous observations of language. Likewise, one cannot imagine a corpus linguist looking into a corpus completely without a preconceived notion of what she is looking for, in other words without a 'probing question' in mind. Corpus linguists do not simply collect data and let the computers churn out analyses.

¹⁹ For example, while Chomsky has always employed transformations (movement, insertion, and deletion), most alternative generative theories reject them. Another split has to do with the status of grammatical relations (subject, object, etc.). While RG and LFG consider them primary notions, others treat them as secondary.

Each of them must work within a linguistic framework, however general, within which the data can be interpreted. Once again, the disparity may not be as wide as it first appears.

In our view, grammar and use are the two sides of a coin and the two approaches focus on one side each. The E-language of a language user is an end product of I-language, and the I-language of a language acquirer is in turn the result of exposure to E-language. The two forms thus interact in profound ways. As aptly pointed out in Her (2005), linguistics is after all a science, and as such it must be at the end a collective endeavor. Thus, a linguist can surely choose to focus on one specific, narrow area of study, or employ one particular scientific methodology and have a meaningful contribution nonetheless. At the present, each side has its loyal followers; one should be encouraged to examine how the two approaches may interact. The existence of one does not necessarily pose a threat to the other, and both may in fact be necessary for a comprehensive view of *language* and *languages*. There is no logical reason why they cannot ‘converge and compensate for each other’.

6. A case study of reconciliation

In this section we first examine the case study of the Mandarin lexical unit of 火勢 *huoshi* ‘fire’, which Shei (2004) uses to illustrate the concept of ‘extended lexical unit’, which he claims ‘may well generate thoughts about treating grammar of human language in an alternative fashion’. One of the converging points among contemporary syntactic theories is indeed the increased role of the lexicon. Sentence structure is generally considered predictable from lexical meanings (e.g. Wasow 1985:204). Certain lexicalist generative theories, for example Word Grammar (cf. e.g. Hudson 1994) and Lexicase (cf. e.g. Starosta 1988), among others, even take the extremist view that the lexicon is the entire grammar and have done without phrase structures or phrase structure rules. It is thus little wonder that some of the corpus linguists have taken up this direction in their research. In short, lexicalism has been around for a long time and, while it may run counter to the derivational syntactic theory advocated by Chomsky, it is by no means a threat to generative linguistics in general.

6.1 Re-examining Shei’s case study

Shei’s case study involves the so-called ‘extended lexical unit’ headed by 火勢 *huoshi* ‘fire flames’. His supporting data consist of 38 tokens of the sentence 火勢一發不可收拾 *huoshi yifa buke shoushi* ‘once fire broke out it could not be contained’

and 10 instances of 火勢延燒_____ *huoshi yanshao* _____ ‘the fire’s momentum continued...’, 5 of which fit in the specific pattern 火勢延燒_____才(被/獲得)控制/撲滅 *huoshi yanshao* _____ *cai (bei/huode) kongzhi/pumie* ‘the fire’s momentum continued to spread _____ before it was contained’.

First of all, even if this case study is convincingly presented and formally formulated it would pose no threat to generative linguistics, let alone when presented in such casual and impressionistic terms: ‘the central word 火勢 heads the unit followed by a structure denoting the consequence of the fire’ (Shei 2004:9). Secondly, one wonders whether Shei would not find it useful to formally formulate this ‘structure denoting the consequence of the fire’ in an established generative syntactic framework. This concept of ‘extended lexical unit’ can also be easily subsumed under the concept of construction, in Construction Grammar, HPSG, or LFG for example. Finally, note the interesting fact that these data are all taken from a corpus of fire reports. The following is an irony that Chomsky once perpetrated:

[I]n an early discussion of Robert Dixon’s claim that *milk* and *white* would be likely to co-occur in British culture, [Chomsky] writes: ‘The high probability of the phrase “white milk” in British English was a fact unknown to me before reading this’. (Smith 1999:203)

Shei’s assertion here seduces one to perpetrate the same kind of irony: The high probability of the term 火勢 *huoshi* ‘fire, fire momentum’ appearing before a structure denoting the consequence of the fire in *fire reports* was an astonishing discovery! After all, these are fire reports. Had a fire not had a certain degree of severity, which demanded the attention of the authority, and not been contained eventually, there would not have been a fire report. A good scientist should strike the right balance between empirical respectability and theoretical coherence. Corpus linguistics and generative linguistics are at the two ends of the scale. Here is a lesson from evolution science:

Most human evolution theories of twenty or fifty years ago are barely worth reading now because, by tying themselves too closely to the physical evidence then available, they aimed too much for empirical respectability at the expense of theoretical coherence. The theories that remain relevant are those derived from fundamental principles of evolutionary biology and commonsense observations about the human mind. (Miller 2000:23)

Shei's case study has paid a hefty price for tying its analysis too closely to the scanty data itself and thus accomplishes little. The reader is better advised to consult the primary literature of corpus linguistics, e.g. Stubbs (2002) and others cited in this current paper, the other three articles in the June 2004 issue of *Concentric*, and also those cited in Shei, for better-illustrated examples.

6.2 A case study of reconciliation

Now we will provide a case study of our own to demonstrate the fundamental difference between corpus linguistics and generative linguistics and how they do not necessarily threaten each other. The case study, adopted from Her (2005), is on 'the rule of the road', the simple fact that automobiles in any country must keep to one side of the road, either left or right. Imagine a corpus linguist and generative linguist are both trying to come up with a comprehensive description of the rule of the road on earth. The most obvious way is by way of listing. The Australian Automobile Association, for example, provides its members applying for an international driver's license a small handbook, where a fairly comprehensive list is available. Corpus linguists should be happy to call this the corpus approach.

(6) The Rule of the Road on Earth (1st formulation):

List:

[Country]	[Side]
Australia	left
China	right
Japan	left
Taiwan	right
USA	right
etc.	

The problem, however, is that within this approach each entry is a stipulation. It is not clear at all whether a third option, besides left and right, is possible. For example, is it possible for an existing or future country to allow both sides, for example driving is to be on the left side after midnight and before noon, on the right side otherwise? Is it possible that driving be restricted to the center of the road? This list does not *in principle* rule out such ridiculous possibilities. A good corpus linguist thus surely should and would go one step further and *induce* a general rule from the (limited or comprehensive) data already collected and come up with the following description instead, which admittedly looks very much like a 'universal grammar' for driving.

(7) The Rule of the Road on Earth (2nd formulation):

Rule: Within a country, keep to either left or right consistently.

List:

[Country]	[Side]
Australia	left
China	right
Japan	left
Taiwan	right
USA	right
etc.	

Now, a generative linguist can just as easily have achieved a similar ‘universal grammar’ like the one above, the only difference being that she has started out assuming such universal principles existed and then looked for evidence. Note she can come up with such a UG after examining data from merely two countries, each with a different ‘typology’, and then test this grammar against further data. In fact, the UG in (7) can be stated precisely within the Principles and Parameters framework (c.f., Chomsky 1981).

(8) The Rule of the Road on Earth (3rd formulation):

Principle: Within any country, drive on x side of the road only.

Parameter: $x = \text{left/right}$

List:

[Country]	[Side]
Australia	left
China	right
Japan	left
Taiwan	right
USA	right
etc.	

The UG in (8) is much like the Head Parameter, which states that a language is either head-initial (e.g. VO) or head-final (e.g. OV). An explicit statement is still needed however for each language, hence the list in (8).

The corpus approach and the generative approach thus do indeed converge and yield the same analysis and reveal the same insight. However, this is probably also the point where the two approaches diverge. The corpus linguist, having pretty much exhausted the data and arrived at a satisfactory analysis, is now quite happy to

conclude the project. The generative linguist, however, due to the constant pursuit of simplicity, is still haunted by the question: is there a simpler solution? The quest now has little to do with data, which have been accounted for already; rather, it is due to theoretical pressure. The generative linguist is likely to be concerned with the disjunction in the parameter, $x = left/right$, and the stipulation of each setting in the list. It will make a far simpler UG if the parameter and the list can both be gotten rid of. In other words, a single universal principle that accounts for everything would be much preferred. An inspiration may come to this generative linguist living in Taipei, as she enters her car: ‘The drive’s seat is on the left side of the car, and yet I drive on the right side of the road.’ Being the good generative linguist that she is, she immediately returns to her lab and takes no time at all to derive the following universal principle:

(9) The Rule of the Road on Earth (4th formulation):

Principle: Drive on the right side of the road if the driver’s side in the automobile is on the left; otherwise, drive on the left.

Further testing of course confirms this far simpler grammar. What an improvement this has made: no more parameter and no more listing! However, after the initial excitement subsides, she soon realizes that she is still not completely happy because the disjunction in (9) due to the left versus right opposition is still a snag. Driving home that evening, after picking up her husband, it dawns on her that she, the driver, is always closer to the center of the road than the front seat passenger. As demonstrated in the following two figures, this is true whether the driver is in Taiwan, which requires driving on the right, or in Australia, which requires driving on the left.

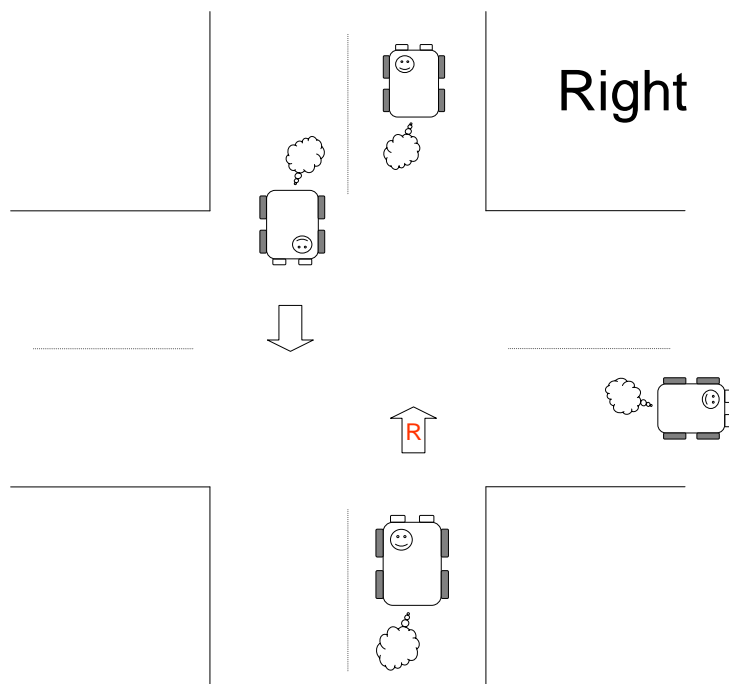


Figure 1. Driving on the right side of the road

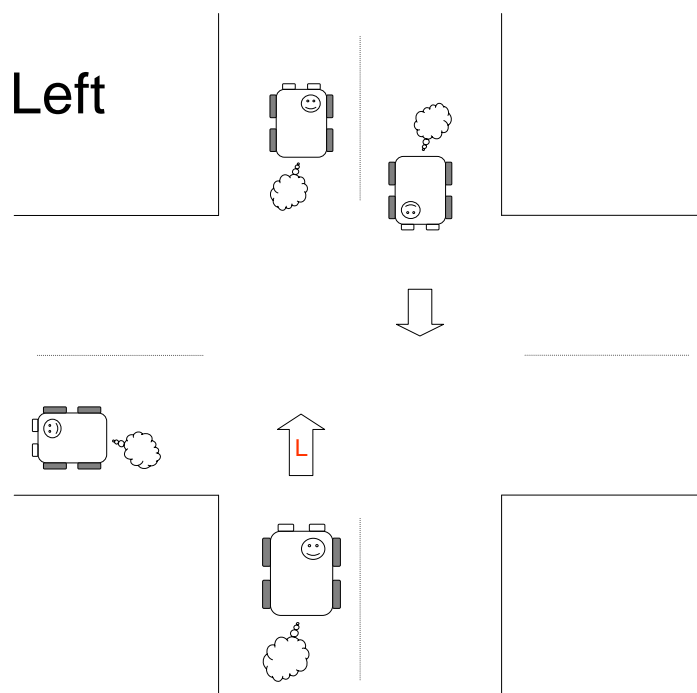


Figure 2. Driving on the left side of the road

The driver and the center of the road are both constants, not disjunctive parameters. The generative linguist is thus able to get rid of the left/right parameter and also the list where each country must require an arbitrary setting, with the following simple universal grammar.

(10) The Rule of the Road on Earth (final formulation):

Principle: The driver must be closer to the center of the road than the front passenger seat is.

The attentive reader with sufficient background in generative syntax must have by now recognized the rather *minimalist* flavor of (10), which is not likely to have emerged without the relentless pursuit for simplicity based on the assumption that such simple universal principles exist. Readers who are familiar with X-bar theories probably have also detected the similarity between (10) and the universal principles proposed by Kayne (1994) regarding phrase structures.

Furthermore, as an anonymous reviewer has keenly observed, a functionalist would certainly not be satisfied with the formal generalization in (10) either and would look further for an explanation that is external to the side of the road. Indeed such a functional explanation has been proposed in Her (2005:61). He contends that the principle in (10) reveals a crucial insight into the functionality of this universal design: being closer to the center of the road affords the driver the best range of vision with the least physical strain. This solution thus also serves as an example of how a formalist analysis and a functionalist account can be satisfactorily reconciled and even complement each other.

However, in order to find universality in seemingly diversified data, one must first look for it and one must look in the right places. The kind of linguistics Shei advocates does not encourage either. The generative paradigm, for better or for worse, emphasizes both. Here is another inspiring example from anthropology:

When anthropologists claim that standards of beauty vary capriciously from one culture to another, they are usually studying the wrong traits in the wrong ways. Individuals of different cultures may like skin of different shades, but they all prefer clean, smooth, unwrinkled skin. Women differ in the exact male height they prefer, but almost always prefer a male taller than themselves. Different ethnic groups may prefer different facial features, but all prefer faces that are symmetrical and averagely shaped for their population. *If you don't look for the universals in human beauty at the right level of description, you will not find them.* (Miller 2000:228, emphasis added)

However, while convinced that (10) is superior to all the previous formulations, we must also concede that (10) is far from being a complete universal description, because exceptions abound. We first give a few examples from the perspective of the automobile. Some countries, Japan for example, allow imported automobiles with the

steering wheel on the wrong side. Still others allow tourists with non-conforming vehicles to cross the border. Myanmar (*a.k.a.* Burma) requires driving on the right, even though most automobiles in this ex-British colony are right-hand-drive. Also, automobiles used exclusive for mail delivery service in many countries, the US and Australia for example, have the driver on the opposite side. From the perspective of the road, (10) also fails for one-way streets, one-lane bridges, and some of the U-turn by-passes that require driving on the opposite side, e.g. the ones along Jianguo South Road under the Jianguo Overpass in Taipei. Thus, while the generative linguist now may rest, a comprehensive description still requires efforts by the corpus linguist. This is why Fillmore has reached the following conclusion:

...the two kinds of linguists need each other. Or better, that the two kinds of linguists, wherever possible, should exist in the same body. (Fillmore 1992:35)

From the broader perspective of language evolution, theoretical or formalist linguistics and functional or corpus linguistics also can, and arguably must, co-exist. Miller (2000:352), in spite of his unequivocal endorsement of the generativist position on language innateness, observes this imbalance:

We have a quandary: the syntax theorists who study grammaticality judgments dominate the conferences on language evolution, while the sociolinguists who study the more evolutionarily important social judgments about speakers will not talk to evolutionary psychologists. ...[S]yntax, morphology, and semantics are the core of human language—but from a Darwinian point of view they are just the incredibly complex design details of a signaling adaptation centered upon social functions and social content.

It is clear where Miller's good taste lies. However, what is equally clear is that a comprehensive understanding of language requires a thorough understanding of language's formal structures as the foundation, only based on which can insights into their social and communicative functions be revealed. After all, in order to understand how and for what purposes the signaling adaptation evolved, don't we need to understand the 'incredibly complex design' of this signaling adaptation first?

7. Conclusion

In this final section we will first address the issues raised in Shei's conclusion, before offering our own concluding remarks. Before stating his conviction that no

future linguists should pursue generative linguistics, Shei first took several parting shots, which do require responses.

7.1 Innateness and language instinct

Shei's first shot is by way of quoting Cowley (2001), which is a direct attack on Pinker's thesis of language instinct:

Referring to Sampson (1997), Cowley (2001) put it somewhat humorously: 'What Sampson, to my judgement (*sic*), successfully demolishes is Pinker's thesis that we possess innate knowledge of the X bars, traces, cases and other representations posited in generativist theory,' and then somewhat and alarmingly to current and future linguistic workers: 'Forty years of intensive research have failed to provide any empirical base for their theorizing.'

It is unclear what exactly Shei himself is against here, however. It certainly cannot be the innateness hypothesis, given his explicit admission that language is innate. He also fails to note that while Cowley (2001) is against the thesis that language is a human instinct, he is equally opposed to Sampson's conservative view that language is a learned skill, a position Shei seems to be in favor of:

To argue that grammar is purely 'cultural' also proves unsatisfactory. In Sampson's terms, indeed, it leads to dualism and/or a belief in a 'haunted universe'. Thus, it implausibly suggests that neither real-time events, development, nor evolution have grammatical consequences. To avoid allowing the biological baby to disappear with the generativist bathwater, I propose we reject Pinker and Sampson's basic shared assumption. (Cowley 2001:88)

While there is no space here to go over the debate over language instinct, we stress once again that Pinker's view is accepted by not only most generative grammarians but also most evolution scientists (Miller 2000:344), as well as scholars in other adjacent fields (Postal 2005:vii). A much more important point, however, is that generative grammar does not hinge upon the concept of language instinct, because, given the fact that human language is infinite, a psychologically adequate grammar must be a generative grammar. Chomsky himself in fact does not endorse Pinker's adaptationist view, which probably surprises many supporters of generative linguistics (Miller 2000:344-345). Rather, he speculates that language may have arisen, rather suddenly, as a complex system due to the interaction of

non-language-specific general properties enabled by a sufficiently large brain (e.g. Chomsky 2002, 2006:173-185 and Lasnik and Uriagereka 2005:27-28). Thus, the current converging point is that human language depends on innate genetic capacities, but the diverging point is exactly how such capacities have come about during the course of evolution. Shei seems to have confused language instinct with innateness, and innateness with generative linguistics.

7.2 Machine translation and linguistic theories

Shei then cites the fact that among the 36 papers in *Readings in Machine Translation*, edited by Nirenburg et al. (2003), only three refer to Chomsky's early works published in 1955, 1957, and 1965. This plus a single random experience of translating his English abstract into French with a web-based machine translation (MT) system and getting an unexpectedly fair result is enough to convince him that: 'Machine translation has its own linguistic models of a theoretical nature to which generative grammar appears to contribute little.' But personal experience is not useful in testing scientific theories (e.g. Miller 2000:428).

It is true that, given MT's intense demand for practicality, theoretical coherence often gives way to practical considerations. In the introduction to the volume mentioned above, Nirenburg (2003:3) writes:

From the 1960s on, MT was, in fact, often used to apply contemporary linguistic theories, but the systems that were directly inspired by a particular linguistic theory were usually seldom comprehensive or broad-coverage. The discrepancy between the needs of MT and the goals and theories in linguistics is real, and the relationships between MT and the would-be primary 'natural' source of inspiration for MT research are still not very close.

But Shei's bias prevented him from noting that *none* of the 36 papers in this book cited any work by functionalist linguists. Other than the fact that Chomsky is cited in three papers in the volume, Johnson et al. (2003) acknowledge that the EUROTRA system owes its formalism to LFG, and Kaplan et al. (2003) describe an MT design by structural correspondence also within LFG, and Landsbergen (2003) discusses an MT system based exclusively on Montague Grammar. That the volume mentions no application of any non-generative linguistic theory is not a trivial fact, but still it is dangerous to base one's support for the generative theory on such observations. If the number of citations is any indication of the applicability of a theory, as Shei clearly implies, then surely Chomsky's theory on language would be far more applicable than

any of its competitors, in any domain of application, as there is no doubt that Chomsky is the most cited linguist ever, but this is hardly the point. For nearly twenty years, the first author of this current paper has worked on a multilingual machine translation system,²⁰ which employs a parsing engine and a transfer program based on LFG and thus can state with confidence that the various generative frameworks have contributed to the general field of computational linguistics or natural language processing, which MT is part of, far more than any other competing approach. So much so, a comparison is hardly justified. While recent trends in machine translation have turned away from rule-based models in favor of probability-based models, pre-analyzed data based on a generative framework, for example in the form of a tree bank, are extensively used.

7.3 Separating psychology from biology

Shei's final shot finally reveals a deep-seated misunderstanding which leads to his fundamental dissatisfaction with the generative theory:

If generative grammar is to become a convincing theory of human language, more concrete evidence must be supplied, such as the finding of a genome which unambiguously moves a *wh*-word, assigns a θ -role, or guards the principle of c-command. But until this is possible, generative grammar remains a "toy theory" of little practical use and it may even divert researchers from the pursuit of knowledge. (Shei 2004:10)

The generative theory, like any other theory of linguistics, merely provides a computational model for language in the mind. Concepts such as *wh*-movement, thematic roles, and c-command need not be exclusive, or perhaps even necessary, to the generative approach to language. Such grammatical concepts, like concepts from other planes, e.g. phonology, discourse, and pragmatics, can be just as useful for corpus linguistics or indeed even functional linguistics. By asking for evidence at the genetic level for such concepts, Shei confuses linguistics with genetics, psychology with biology. In 1998, Robert Plomin and his team identified the first specific gene, IGF2G, on chromosome 6, associated with high intelligence. By Shei's logic, prior to Plomin's discovery all theories attempting to account for human intelligence were wastes of time. Note that Plomin was also quick to point out that in fact numerous

²⁰ For this LFG-based MT design, see Her et al. (1991, 1994). The system now covers the following language pairs: English→Chinese, English→Japanese, English↔Korean, English↔Turkish, and English↔Arabic, plus prototypes of a few other language pairs. See <http://www.apptek.com> for further details.

genes contribute to intelligence, either negatively or positively (Plomin and DeFries 1998). Likewise, there are probably hundreds or even thousands of genes responsible for language (Miller 2000:23). But without a theory of language innateness, no scientist will even start looking for them. The current consensus in all language-related fields is that language is part of human cognitive endowment and thus an object of study in cognitive psychology, which rose in the 1970s partly due to generative linguistic views. Long gone are the days of Skinner's behaviorist view that reduces language to associated conditions.

7.4 Competition and cooperation

As we have said in the introduction, we do not presume to be able to prove the plausibility of the generative approach to everyone. It is too insurmountable a task. The primary literature would do a far better job. Rather, we have taken the competing rival functionalist linguistics and corpus linguistics for granted and set out to demonstrate that generative linguistics is likewise a worthy research program and that generative linguistics and functionalist linguistics or corpus linguistics can indeed converge and complement one another. Chomsky, contrary to common (mis)perception, *can* be humble at times:

No one familiar with the field has any illusion that the horizons of inquiry are even visible, let alone at hand. (Chomsky 2006:182)

Science is a bit like the joke about the drunk who is looking under a lamppost for a key that he has lost on the other side of the street, because that's where the light is. It has no other choice. (Chomsky, quoted in Barsky (1997, Chp 3))

Only in this case the generative linguists and the corpus linguists, as we have argued, are under the same lamppost. The generative linguists are looking for traces that may lead to the horizons or even the key to UG itself, firmly convinced of its existence, while the corpus linguists are far more practical and simply examine whatever they can find, not particularly concerned whether what they find may or may not lead to a key. The two kinds of linguists can work together and share what they find within their respective approaches, or they can spend their energy trying to topple each other. We hope we have demonstrated that, given the current state of linguistic science, both competition and cooperation between the two approaches are beneficial to the field.

References

- Aarts, Bas. 2000. Corpus linguistics, Chomsky and Fuzzy Tree Fragments. *Corpus Linguistics and Linguistic Theory*, ed. by Christian Mair, and Marianne Hundt, 5-13. Amsterdam: Rodopi.
- Baker, Mark. 1988. *Incorporation: A Theory of Grammatical Function Changing*. Chicago: University of Chicago Press.
- Barsky, Robert F. 1997. *Noam Chomsky: A Life of Dissent*. Cambridge, MA: MIT Press.
- Bates, Elizabeth. 2003. Natura e cultura nel linguaggio [On the nature and culture of language]. *Frontiere della biologia* [Frontiers of biology]. *Il cervello di Homo sapiens* [*The brain of homo sapiens*], ed. By R. Levi-Montalcini, D. Baltimore, R. Dulbecco, and F. Jacob (Series Eds.) and E. Bizzi, P. Calissano, and V. Volterra, 241-265. Rome: Istituto della Enciclopedia Italiana fondata da Giovanni Treccani S.p.A.
- Bever, Thomas, and Brian McElree. 1988. Empty categories access their antecedents during comprehension. *Linguistic Inquiry* 19:35-43.
- Biber, Douglas, Susan Conrad, and Randi Reppen. 1998. *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press.
- Bickerton, Derek. 2007. Language evolution: A brief guide for linguists. *Lingua* 117:510-526.
- Bresnan, Joan, and Judith Aissen. 2002. Functionality and optimality: Objections and refutations. *Natural Language and Linguistic Theory* 20:81-95.
- Brodmann, Korbinian. 1909. *Vergleichende Lokalisationslehre der Grosshirnrinde in ihren Principien, dargestellt auf grund des Zellenbaues*. Leipzig: Johann Ambrosius Barth Verlag. (English translation by Laurence J. Garey, 1994: *Brodmann's Localisation in the Cerebral Cortex*, London: Smith-Gordon, new edition 1999, London: Imperial College Press.)
- Bybee, Joan, and Paul Hopper (eds.). 2001. *Frequency and the Emergence of Linguistic Structure*. Amsterdam and Philadelphia: John Benjamins.
- Cavalli-Sforza, L. L., A. Piazza, P. Menozzi, and J. Mountain. 1988. Reconstruction of human evolution: Bringing together genetic, archaeological and linguistic data. *Proceedings of the National Academy of Science* 85, 6002-6006.
- Cheung, Hintat. 2003. Memory capacity in school-age Mandarin-speaking children with specific language impairment. *Taiwan Journal of Linguistics* 1:111-120.
- Chomsky, Noam. 1955. *Logical Structure of Linguistic Theory*. New York: Plenum.
- Chomsky, Noam. 1957. *Syntactic Structures*. The Hague: Mouton.

- Chomsky Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, Mass.: MIT Press.
- Chomsky, Noam. 1975. *Reflections on Language*. New York: Pantheon.
- Chomsky, Noam. 1981. *Lectures on Government and Binding*. Berlin and New York: Mouton de Gruyter.
- Chomsky, Noam. 1986. *Knowledge of Language*. New York: Praeger.
- Chomsky, Noam. 2000. *The Architecture of Language*, ed. by Nirmalanshu Mukherji, Bibudhendra Narayan Patnaik, and Rama Kant Agnihotri. New York: Oxford University Press. (Excerpts available online at <http://www.chomsky.info/books/architecture01.htm>)
- Chomsky, Noam. 2002. *On Nature and Language*, ed. by Adriana Belletti, and Luigi Rizzi. Cambridge: Cambridge University Press.
- Chomsky, Noam. 2006. *Knowledge of Language*. New York: Cambridge University Press.
- Cowley, S.J. 2001. The baby, the bathwater and the ‘language instinct’ debate. *Language Sciences* 23:69-91.
- Crystal, David. 1997. *The Cambridge Encyclopedia of Language*. Cambridge: Cambridge University Press.
- Dingwall, William O. 1998. The biological bases of human communicative behavior. *Psycholinguistics* (2nd edition), ed. by J. B. Gleason, and N. B. Ratner, 51-105. Orlando, FL: Harcourt College Publishers.
- Everett, Daniel L. 2005. Biology and language: A consideration of alternatives (a review article of Stephen R. Anderson, and David W. Lightfoot, *The Language Organ: Linguistics as Cognitive Physiology*, 2002, Cambridge: Cambridge University Press and Talmy Givón, *Bio-linguistics: The Santa Barbara Lectures*, 2002, John Benjamins: Amsterdam). *Journal of Linguistics* 41:157-175.
- Fillmore, Charles, Paul Kay, and Mary O’Connor. 1988. Regularity and idiomaticity in grammatical constructions: The case of “let alone”. *Language* 64:501-538.
- Fillmore, Charles. 1992. “Corpus linguistics” or “computer-aided armchair linguistics”. *Directions in Corpus Linguistics: Proceedings of Nobel Symposium 82 Stockholm*, August 4-8, 1991, ed. by Jan Svartvik, 35-60. Berlin and New York: Mouton de Gruyter.
- Givón, Talmy. 2002. *Bio-linguistics: The Santa Barbara Lectures*. Amsterdam: John Benjamins.
- Goldberg, Adele. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.
- Goldberg, Adele, and Ray Jackendoff. 2004. The English resultative as a family of constructions. *Language* 80:532-568.

- Grodzinsky, Y. 2000. The neurology of syntax: Language use without Broca's area. *Behavioral and Brain Sciences* 23:1-71.
- Her, One-Soon, Dan Higinbotham, and Joseph Pentheroudakis. 1991. An LFG-based machine translation system. *Computer Processing of Chinese and Oriental Languages* 5:285-297.
- Her, One-Soon, Dan Higinbotham, and Joseph Pentheroudakis. 1994. Lexical and idiomatic transfer in machine translation: An LFG approach. *Research in Humanities Computing*, ed. by Don Ross, and Dan Brink, 200-216. Oxford: Oxford University Press.
- Her, One-Soon. 1991. *Grammatical Functions and Verb Subcategorization in Mandarin Chinese*. Taipei: Crane Publishing.
- Her, One-Soon. 2005. 'Driving' towards simplicity: In (partial) defense of the 'armchair linguist'. *Journal of Language and Linguistics* 4:56-73. Available online at: http://www.shakespeare.uk.net/journal/4_1/onesoonher.htm
- Huang, James C.-T. 1982. *Logical Relations in Chinese and the Theory of Grammar*. Ph.D. dissertation, MIT.
- Hudson, Richard. 1994. Word Grammar. *The Encyclopedia of Language and Linguistics*, ed. by R. E. Asher, and J. M. Y. Simpson, 4990-4993. Oxford: Pergamon.
- Jenkins, Lyle. 2000. *Biolinguistics: Exploring the Biology of Language*. Cambridge: Cambridge University Press
- Ji, Sungchul. 1997. Isomorphism between cell and human languages: Molecular biological, bioinformatic and linguistic implications. *BioSystems* 44:17-39.
- Johnson, George. 1995. Chimp talk debate: Is it really language? *The New York Times*, June 6, 1995, Tuesday, Late edition, section C, page 1, column 1, Science Desk. Available online at: <http://www.santafe.edu/~johnson/articles.chimp.html>
- Johnson, Rod, Maghi King, and Louis des Tombe. 2003. EUROTRA: Computational techniques. *Readings in Machine Translation*, ed. by Nirenburg et al., 345-350. Cambridge, Mass.: MIT Press.
- Kaplan, Ronald, Klaus Netter, Jurgen Wedekind, and Annie Zaenen. 2003. Translation by structural correspondences. *Readings in Machine Translation*, ed. by Nirenburg et al., 263-272. Cambridge, Mass.: MIT Press.
- Kay, Paul, and Charles Fillmore. 1999. Grammatical constructions and linguistic generalizations: The what's X doing Y? construction. *Language* 75:1-33.
- Kayne, Richard. 1994. *The Antisymmetry of Syntax*. Cambridge, Mass: MIT Press.
- Kennedy, Graeme. 1998. *An Introduction to Corpus Linguistics*. London: Longman.

- Landsbergen, Jan. 2003. Montague Grammar and machine translation. *Readings in Machine Translation*, ed. by Nirenburg et al., 239-254. Cambridge, Mass.: MIT Press.
- Lasnik, Howard, Juan Uriagereka, and Cedric Boeckx. 2005. *A Course in Minimalist Syntax*. Oxford: Blackwell.
- Leech, Geoffrey. 1992. Corpora and theories of linguistic performance, Directions in Corpus Linguistics. *Proceedings of Nobel Symposium 82*, August 4-8, 1991, ed. by Jan Svartvik, 105-122. Trends in Linguistics Studies and Monographs 65. Berlin: Mouton.
- Mair, Christian, and Marianne Hundt (eds.). 2000. *Corpus Linguistics and Linguistic Theory* (Language and Computers: Studies in Practical Linguistics, No. 33). Amsterdam and Atlanta: Rodopi.
- McElree, Brian, and Thomas Bever. 1989. The psychological reality of linguistically defined gaps. *Journal of Psycholinguistic Research* 18:21-35.
- Meyer, Charles. 2002. *English Corpus Linguistics: An Introduction*. Cambridge: Cambridge University Press.
- Miller, Geoffrey. 2000. *The Mating Mind: How Sexual Choice Shaped the Evolution of Human Nature*. New York: Doubleday.
- Moro, A., M. Tettamanti, D. Perani, C. Donati, S. F. Cappa, and F. Fazio. 2001. Syntax and the brain: Disentangling grammar by selective anomalies. *NeuroImage* 13:110-118.
- Müller, R.-A., N. Kleinhans, and E. Courchesne. 2003. Linguistic theory and neuroimaging evidence: An fMRI study of Broca's area in lexical semantics. *Neuropsychologia* 41:1199-1207.
- Newmeyer, Frederick. 1991. Functional explanation in linguistics and the origins of language. *Language and Communication* 11:3-28.
- Newmeyer, Frederick. 2003. Grammar is Grammar and Usage is Usage. *Language* 79:682-707.
- Nirenburg, Sergei, Harold Somers, and Yorick Wilks (eds.). 2003. *Readings in Machine Translation*. Cambridge, Mass.: MIT Press.
- Nirenburg, Sergei. 2003. Introduction. *Readings in Machine Translation*, ed. by Nirenburg et al., xi-1. Cambridge, Mass.: MIT Press.
- Pinker, Steven. 1994. *The Language Instinct: How the Mind Creates Language*. New York: William Morrow.
- Pinker, Steven, and Paul Bloom. 1990. Natural language and natural selection. *Behavioral and Brain Sciences* 13:707-784.
- Plomin, Robert, and John DeFries. 1998. The genetics of cognitive abilities and disabilities. *Scientific American* 278:40-47.

- Postal, Paul. 2005. Forward to *The 'Language Instinct' Debate*, Geoffrey Sampson. London and New York: Continuum International.
- Ratner, Bernstein, J. Berko Gleason. and Bhuvana Narasimhan. 1998. Introduction to psycholinguistics: What do language users know? *Psycholinguistics*, ed. by J. Berko Gleason, and N. Bernstein Ratner, 1-40. 2nd edition. Orlando, FL: Harcourt Brace.
- Rice, Mabel. 2002. A unified model of specific and general language delay: Grammatical tense as a clinical marker of unexpected variation. *Language Competence across Populations: Toward a Definition of Specific Language Impairment*, ed. by Y. Levy, and J. Schaeffer, 63-95. Mahwah, NJ: Lawrence Erlbaum Associates.
- Ritter, Nancy. 2002. Introduction. *Linguistic Review* 19:1-7.
- Sampson, Geoffrey. 1997. *Educating Eve: The 'Language Instinct' Debate*. London: Cassell.
- Sampson, Geoffrey. 2005. *The 'Language Instinct' Debate*. London and New York: Continuum International.
- Savage-Rumbaugh, Sue, and R. Lewin. 1994. *Kanzi: The Ape at the Brink of the Human Mind*. New York: Wiley.
- Savage-Rumbaugh, Sue. 1999. Ape language: Between a rock and a hard place. *The Origins of Language: What Nonhuman Primates Can Tell Us*, ed. by Barbara J. King, 115-189. Santa Fe: School of American Research Press.
- Schutze, Carson T. 1996. *The Empirical Base of Linguistics: Grammaticality Judgments and Linguistic Methodology*. Chicago: University of Chicago Press.
- Sells, Peter. 1985. *Lectures on Contemporary Syntactic Theories*. Stanford University, Stanford, CA.: CSLI Publications.
- Sells, Peter. (ed.). 2001. *Formal and Empirical Issues in Optimality Theoretic Syntax*. Stanford, CA: CSLI Publications.
- Shei, Chris Chi-Chiang. 2004. Corpus and grammar: What it isn't. *Concentric: Studies in Linguistics* 30:1-18
- Simpson, Joy, and Mabel Rice. 2004. *Top 10 Things You Should Know about Children with Specific Language Impairment*, article posted on <http://merrill.ku.edu/IntheKnow/sciencearticles/SLIfacts.html>
- Smith, Neil, and Ianthi Tsimpli. 1995. *The Mind of A Savant: Language Learning and Modularity*. Oxford: Blackwell.
- Smith, Neil. 1999. *Chomsky: Ideas and Ideals*. Cambridge: Cambridge University Press.
- Snow, Catherine E. 1996. Toward a rational empiricism: Why interactionism is not behaviourism any more than biology is genetics. *Toward A Genetics of Language*,

- ed. by Mabel L. Rice, 377-396. Mahwah, NJ: Lawrence Erlbaum Associates.
- Starosta, Stanley. 1987. *Generative Syntax: The Hypothetico-Deductive Science of Language*. Taipei: The Crane Publishing Company.
- Sarosta, Stanley. 1988. *The Case for Lexicase: An Outline of Lexicase Grammatical Theory*. London: Pinter Publishers Limited.
- Stubbs, Michael. 2002. *Words and Phrases: Corpus Studies of Lexical Semantics*. Oxford: Blackwell Publishing.
- Svartvik, Jan (ed.). 1992. Directions in Corpus Linguistics. *Proceedings of Nobel Symposium 82*, August 4-8, 1991. Trends in Linguistics Studies and Monographs 65. Berlin: Mouton.
- Tsai, Mei-chih. 1998. *Gaoxing* and *Kuaile*—Two types of state of happiness, paper presented at the European Association of Chinese Studies EACS 1998 Conference, Edinburgh, September 10-13.
- Tsai, Mei-Chih, Chu-Ren Huang, Keh-Jiann Chen, and Kathleen Ahrens. 1998. Towards a representation of verbal semantics: An approach based on near-synonyms. *Computational Linguistics and Chinese Language Processing* 3: 61-74.
- Ullman, Michael, Suzanne Corkin, Marie Coppola, Gregory Hickok, John H. Growdon, Walter J. Koroshetz, and Steven Pinker. 1997. A neural dissociation within language: Evidence that the mental dictionary is part of declarative memory, and that grammatical rules are processed by the procedural system. *Journal of Cognitive Neuroscience* 9:266-276.
- Wasow, Thomas. 1985. Postscript. *Lectures on Contemporary Syntactic Theories*, ed. by Peter Sells, 193-205. Stanford, CA.: CSLI Publications.

[Received 25 July 2006; revised 21 December 2006; accepted 29 January 2007]

Graduate Institute of Linguistics
National Chengchi University
Taipei, TAIWAN
One-Soon Her: onesoon@gmail.com
I-Ping Wan: ipwan@nccu.edu.tw

再論語料庫及語法的本質

何萬順 萬依萍

國立政治大學

Shei (2004) 最重要的兩項主張是：一、心理語言學與生物學研究的相關證據顯示衍生語法學是不可信的，二、詞彙的重要性很可能超過句法；因此語料庫語言學在語法研究上才是較為可行的。本文是對該文的這些論點一一做出回應。我們首先釐清衍生語法學的部分觀點及其現況，並且指出該文論證之不足。本文所提出的相對主張是：一、現有的證據足以顯示衍生語法學是一個值得探索的科學理論，二、衍生語言學與功能語言學或語料庫語言學並不是必然衝突的；本文最終的目的在於證明若要瞭解人類的個別語言以及普遍語法，這兩種不同的理論及其研究方法都是必要的。

關鍵詞：衍生語法學、功能語言學、語料庫語言學、心理語言學