Assessing Grammar
THE CAMBRIDGE LANGUAGE ASSESSMENT SERIES

Series editors: J. Charles Alderson and Lyle E. Bachman

In this series:

Assessing Vocabulary by John Read
Assessing Reading by J. Charles Alderson
Assessing Language for Specific Purposes by Dan Douglas
Assessing Writing by Sara Cushing Weigle
Assessing Listening by Gary Buck
Assessing Speaking by Sari Luoma
Statistical Analyses for Language Assessment by Lyle E. Bachman
Assessing Grammar

James E. Purpura
Contents

Series editors’ preface xi
Acknowledgments xi

1 Differing notions of ‘grammar’ for assessment 1
2 Research on L2 grammar teaching, learning and assessment 24
3 The role of grammar in models of communicative language ability 49
4 Towards a definition of grammatical ability 83
5 Designing test tasks to measure L2 grammatical ability 100
6 Developing tests to measure L2 grammatical ability 146
7 Illustrative tests of grammatical ability 181
8 Learning-oriented assessments of grammatical ability 212
9 Challenges and new directions in assessing grammatical ability 251

References 275
Index 291
CHAPTER ONE

Differing notions of ‘grammar’ for assessment

Introduction

The study of grammar has had a long and important role in the history of second language and foreign language teaching. For centuries, to learn another language, or what I will refer to generically as an L2, meant to know the grammatical structures of that language and to cite prescriptions for its use. Grammar was used to mean the analysis of a language system, and the study of grammar was not just considered an essential feature of language learning, but was thought to be sufficient for learners to actually acquire another language (Rutherford, 1988). Grammar in and of itself was deemed to be worthy of study – to the extent that in the Middle Ages in Europe, it was thought to be the foundation of all knowledge and the gateway to sacred and secular understanding (Hillocks and Smith, 1991). Thus, the central role of grammar in language teaching remained relatively uncontested until the late twentieth century. Even a few decades ago, it would have been hard to imagine language instruction without immediately thinking of grammar.

While the central role of grammar in the language curriculum has remained unquestioned until recent times, grammar pedagogy has unsurprisingly been the source of much debate. For example, some language educators have argued that foreign languages are best learned deductively, where students are asked to memorize and recite definitions, rules, examples and exceptions. In this approach, the teaching of language obviously involved the transmission of grammar rules from teacher to student, and to know a language meant to know the intricacies
of its grammatical system and to recite its rules. Other language educators have maintained that language learning is best achieved inductively. In this approach, students are presented with examples of the target language and led to discover its underlying organizational principles in order to be able to formulate a formal set of rules and prescriptions. To know an L2 here meant to identify and describe the rules of the language system based on an analysis of texts. Still other more traditional language teachers have claimed that the best way to learn an L2 is to study its grammar so that the language could be translated from one language to another. Based upon the assumption that all languages are similar and that Latin could be used as a model for analysis, this practice gave rise to the first grammars for foreign-language learners and to the grammar-translation approach to language learning (Rutherford, 1987), still used in many classrooms around the world today. In this approach, knowledge of a language involves the ability to read and render an accurate translation. In each cited instance of language teaching, grammar has remained the unquestioned focus, and knowledge of the grammar is viewed as a set of rules. Similarly, the assessment of grammatical knowledge is carried out by having students recite rules, by having them analyze texts and state the rules, or by having them translate texts. In short, grammatical assessment was closely aligned with the goals of instruction and, until recent times, was hardly a topic of concern.

It was not until the late twentieth century that the central role of grammar in language teaching was seriously questioned. In reaction to the grammar-translation approach that had become more about learning a set of abstract linguistic rules than about learning to use a language for some communicative purpose, some language teachers began to seek alternative approaches to language teaching based on what students could ‘do’ with the language. These teachers insisted that the grammar should not only be learned, but also applied to some linguistic or communicative purpose. They recommended that grammatical analysis be accompanied by application, where students are asked to answer questions, write illustrative examples, combine sentences, correct errors, write paragraphs and so forth. To know a language meant to be able to apply the rules – an approach relatively similar to what is done in many classrooms today. In this approach, knowledge of grammar was assessed by having students apply rules to language in some linguistic context.

Other language teachers have been more vehement in their attempt to de-emphasize the role of grammar in language teaching. They believed that foreign languages were best learned in the same way that children
learn their native language, through sustained exposure to the language and through interaction. This belief gave rise to the ‘natural method’, the ‘direct method’, and, ultimately, to the ‘natural approach’ to language acquisition (Krashen and Terrell, 1983). Although these language-teaching methods differed in terms of whether first and second language acquisition were assumed to involve identical processes, these methods made little or no provision for the formal instruction of grammar, and students were left to their own devices to identify and learn the rules. In these approaches, grammar was no longer seen exclusively as a set of grammatical abstractions to be recited, but rather as a set of rules to be internalized and used for communication. To know a language meant to be able to use it for some real-life purpose, and the assessment of grammatical knowledge was based on tasks requiring students to demonstrate their ability to communicate in speaking or writing.

Most of the early debates about language teaching have now been resolved; however, others continue to generate discussion. For example, most language teachers nowadays would no longer expect their students to devote too much time to describing and analyzing language systems, to translating texts or to learning a language solely for access to its literature; rather, they would want their students to learn the language for some communicative purpose. In other words, the primary goal of language learning today is to foster communicative competence, or the ability to communicate effectively and spontaneously in real-life settings. Language teachers today would not deny that grammatical competence is an integral part of communicative language ability, but most would maintain that grammar should be viewed as an indispensable resource for effective communication and not, except under special circumstances, an object of study in itself. Current teaching controversies revolve around the role, if any, that grammar instruction should play in the language classroom and the degree to which the grammatical system of a language can be acquired through instruction. These questions have, since the 1980s, produced an explosion of empirical research, which is of critical importance to language teachers. Given the significance of this literature for teachers and the critical role that grammatical assessment plays in this research, I will examine this literature in some detail later on in this chapter.

In summary, language teachers have always acknowledged the inextricable link between teaching and testing, and accordingly have always assessed their students’ knowledge of grammar. In other words, the assessment of grammatical ability is nothing new. What has changed over
time is what teachers have chosen to assess under the title of ‘grammar’ and the ways in which these assessments have been carried out. For example, at one point in time, knowledge of grammar was assessed through the ability to recite rules; at another, through the ability to extrapolate a rule from samples of the target language; and at yet another, knowledge of grammar was tested through the ability to provide an accurate translation. Currently, knowledge of grammar might be inferred from the ability to select a grammatically correct answer from several options on a multiple-choice test, to supply a grammatically accurate word or phrase in a paragraph or dialogue, to construct grammatically appropriate sentences, or to provide judgments regarding the grammaticality of an utterance. In many assessment contexts today, knowledge of grammar may be inferred from the ability to use grammar correctly while reading, writing, listening to or speaking the L2 – a practice based on the assumption that all instances of language use invoke the same fundamental working knowledge of grammar and that a lack of grammatical knowledge can severely limit what is understood or produced in communication. In short, language educators have defined and assessed grammatical knowledge in many different ways over the years as the notion of what it means to ‘know’ the grammar of a language has evolved and instructional practices have changed.

What is striking, however, in the long-standing debate on grammar and its role in language learning is the relative absence of discussion of how ‘best’ to assess grammatical knowledge or how to determine if grammatical knowledge has been acquired. Even with the sudden increase of research since the mid-1980s on the teaching and learning of grammar, there still remains a surprising lack of consensus on (1) what constitutes grammatical knowledge, (2) what type of assessment tasks might best allow teachers and testers to infer that grammatical knowledge has been acquired and (3) how to design tasks that elicit grammatical knowledge from students for some specific assessment purpose, while at the same time providing reliable and valid measures of performance. In other words, there is a glaring lack of information available on how the assessment of grammatical ability might be carried out, and how the choices we make in the assessment of grammatical ability might influence the inferences we make about our students’ knowledge of grammar, the decisions we make on their behalf and their ultimate development.

In this first chapter, I hope to add some clarity to these issues by describing what language educators generally mean when they talk about ‘grammar’. I will show the influences of linguistic theory on differ-
ent conceptualizations of ‘grammar’ in L2 educational contexts and will demonstrate how these different descriptions of grammar have broadened our understanding of how language is organized. I will argue that it is important for teachers to have a solid understanding of the grammatical resources of language so that instruction and assessment can be tailored to a variety of educational contexts.

What is meant by ‘grammar’ in theories of language?

Grammar and linguistics

Before attempting to define what it means to ‘know’ grammar or to be able to ‘use’ it to communicate in second or foreign language contexts, we first need to discuss what is commonly meant by ‘grammar’. This is important given the different definitions and conceptualizations of grammar that have been proposed over the years, and the diverse ways in which these notions of grammars have influenced L2 educators.

When most language teachers, second language acquisition (SLA) researchers and language testers think of ‘grammar’, they call to mind one of the many paradigms (e.g., ‘traditional grammar’ or ‘universal grammar’) available for the study and analysis of language. Such linguistic grammars are typically derived from data taken from native speakers and minimally constructed to describe well-formed utterances within an individual framework. These grammars strive for internal consistency and are mainly accessible to those who have been trained in that particular paradigm.

Since the 1950s, there have been many such linguistic theories – too numerous to list here – that have been proposed to explain language phenomena. Many of these theories have helped shape how L2 educators currently define grammar in educational contexts. Although it is beyond the purview of this book to provide a comprehensive review of these theories, it is, nonetheless, helpful to mention a few, considering both the impact they have had on L2 education and the role they play in helping define grammar for assessment purposes.

Generally speaking, most linguists have embraced one of two general perspectives to describe linguistic phenomena. Either they take a syntactocentric perspective of language, where syntax, or the way in which words are arranged in a sentence, is the central feature to be observed and analyzed; or they adopt a communication perspective of language,
where the observational and analytic emphasis is on how language is used to convey meaning (VanValin and LaPolla, 1997). I will use these two perspectives to classify some of the more influential grammatical paradigms in our field.

In the syntactocentric view of language, formal grammar is defined as a systematic way of accounting for and predicting an ‘ideal’ speaker’s or hearer’s knowledge of the language. This is done by a set of rules or ‘principles’ that can be used to generate all well-formed or grammatical utterances in the language. This approach typically examines sounds that are combined to form words, words that are put together to form phrases, phrases combined to form clauses, and clauses assembled to form sentences. In other words, this approach is predominantly concerned with the structure of clauses and sentences, leaving the literal meaning and contextual use of these forms to other approaches (i.e., to the fields of semantics and pragmatics). To illustrate, consider the following sentence:

(1.1) Reggio and Messina were taken to the vet’s this morning.

Some formal grammarians would explain this passive voice sentence by comparing it with its active voice counterpart – *[someone] took Reggio and Messina to the vet’s this morning*. They would then derive a number of rules – for changing the past to the past passive (took → were taken), for moving the patient of the action (Reggio and Messina) to the subject position and for deleting the agent (by someone). They would also devise rules for pronunciation and spelling. Some formal grammarians might even explain this sentence by comparing it to a number of ungrammatical passive sentences.

Syntactocentric theories of language have provided L2 educators with a wealth of information about grammatical forms and the rules that govern them. In fact, most classroom language teachers draw extensively on this information as a basis for syllabus design, materials preparation, instruction and classroom assessment. These theories have also informed L2 teachers and testers in their efforts to identify linguistic content for tests so that more general inferences about language ability can be made.

The second general approach to describing language is through an analysis of communication. In this perspective, the structural description of the language is not the primary object of concern; rather, language is viewed as a system of communication, where a speaker or writer uses
grammatical forms to convey a number of meanings. In the communication perspective, grammar is treated as one of many resources for accomplishing something with language, and grammarians describe both what the linguistic forms are for and how they are used to create meaning within and beyond the sentence. In other words, while the choice of the right grammatical form and the most appropriate lexical item is important, this perspective focuses more on the overall message being communicated and the interpretations that this message might invoke.

Grammarians with a communicative view of language might explain the passive voice sentence in 1.1 in a very different way. They would first take note of the structural features of the passive voice, just as the formal grammarians did. For example, they would compare the following sentences structurally.

(1.1) Reggio and Messina were taken to the vet’s this morning [by someone].
(1.2) [someone] took Reggio and Messina to the vet’s this morning.

However, they would also be interested in the features of the context that required the speaker to choose the passive over the active voice in the first place. In other words, what was the communicative need for the passive? What was the speaker or writer trying to communicate by its use? From a communication perspective, they might determine that the speaker wished to shift the communicative focus from the actors or agent in the sentence (the person who took the cats to the vet) to the recipients of the action (Reggio and Messina). This highlights the fact that Reggio and Messina were taken to the vet’s – since as cats they could not go there by themselves. Thus, the patient of the action (Reggio and Messina) becomes the grammatical subject of the sentence rather than the object.

Of equal interest would be the features of the context that allowed the agent to be omitted since we never learn who actually took Reggio and Messina to the vet’s. Given more contextual information, we could most likely infer this; however, in a single isolated, ‘discrete’ utterance, this information is not available. Furthermore, as the agent in this sentence seemed irrelevant, it went unexpressed. This may also be because the agent is unknown, but it is more likely the case that the agent is known from the context and repeating it would have been redundant. The communication perspective of language, therefore, attempts to examine the relationship between the grammatical forms we use and the conceptual
meanings we wish to express, given the context in which the utterances were situated. Like the syntactocentric perspective, this perspective has much to offer the L2 educator, especially when it comes to using grammar as a resource for communication.

These two views of linguistic analysis have been instrumental in determining how grammar has been conceptualized in L2 classrooms in recent years. They have also influenced definitions of L2 grammar for assessment purposes. I will now provide a brief overview of some of the more influential linguistic theories that typify the syntactocentric and communicative views of language.

Form-based perspectives of language

Several syntactocentric, or form-based, theories of language have provided grammatical insights to L2 teachers. I will describe three: traditional grammar, structural linguistics and transformational-generative grammar.

One of the oldest theories to describe the structure of language is traditional grammar. Originally based on the study of Latin and Greek, traditional grammar drew on data from literary texts to provide rich and lengthy descriptions of linguistic form. Unlike some other syntactocentric theories, traditional grammar also revealed the linguistic meanings of these forms and provided information on their usage in a sentence (Celce-Murcia and Larsen-Freeman, 1999). Traditional grammar supplied an extensive set of prescriptive rules along with the exceptions. A typical rule in a traditional English grammar might be:

The first-person singular of the present tense verb ‘to be’ is ‘I am’. ‘Am’ is used with ‘I’ in all cases, except in first-person singular negative tag and yes/no questions, which are contracted. In this case, the verb ‘are’ is used instead of ‘am’. For example, ‘I’m in a real bind, aren’t I?’ or ‘Aren’t I trying my best?’

Traditional grammar has been criticized for its inability to provide descriptions of the language that could adequately incorporate the exceptions into the framework and for its lack of generalizability to other languages. In other words, traditional grammar postulated a separate, uniquely language-specific set of rules or ‘parameters’ for every language. In spite of these shortcomings as a form of linguistic analysis, traditional grammar has had an enormous impact on L2 teachers and testers.
throughout the centuries, and many L2 educators continue to find it a valuable source of information.

Another influential theory of linguistic analysis grew out of a concerted effort by linguists in the United States both to teach English to Native American Indians and to learn the indigenous American languages so that they could be documented and preserved. However, as these languages in the early twentieth century had no written alphabet and as the native speakers were unable to describe the languages, linguists departed from the long tradition of comparing English to Latin and began to collect samples of the target languages with the goal of providing a description of its phonology (i.e., its sound system), its morphology (i.e., the study of minimal units of meaning or grammatical function such as in *untrue* → *un* + true or *walked* → *walk* + *ed*) and its syntax (Chastain, 1976). This work ultimately gave rise to descriptive or structural linguistics.

Structural grammars, associated with linguists such as Bloomfield (1933) and Fries (1940), offered a fairly rigorous method for describing the structure of a language in terms of both its morphology and its syntax. In these grammars each word in a given sentence is categorized according to how it is used, and the ‘patterns’ or ‘structures’ are said to constitute a unique system for that language. Figure 1.1 shows how a structural grammar might analyze statements and yes/no questions in English.

![Figure 1.1 Structural analysis of statements and yes/no questions in English](image-url)

Unlike traditional grammars, structural grammars are not based on a set of prescriptive rules. Rather, they seek to describe the language as it appears with a strict focus on grammatical form. Although descriptive linguistics has provided numerous insights into the structure of languages, it downplayed the semantic aspects of grammar, and provided little information on how linguistic forms are used in context. Nonetheless, many L2 educators continue to consider this theory a valuable resource for use in syllabus design, grammar teaching and assessment.
Probably the best-known syntactocentric theory is Chomsky’s (1965) transformational-generative grammar and its later, broader instantiation, universal grammar (UG). Unlike the traditional or structural grammars that aim to describe one particular language, transformational-generative grammar endeavored to provide a ‘universal’ description of language behavior revealing the internal linguistic system for which all humans are predisposed (Radford, 1988). Transformational-generative grammar claims that the underlying properties of any individual language system can be uncovered by means of a detailed, sentence-level analysis. In this regard, Chomsky proposed a set of phrase-structure rules that describe the underlying structures of all languages. These phrase-structure rules join with lexical items to offer a semantic representation to the rules. Following this, a series of ‘transformation’ rules are applied to the basic structure to add, delete, move or substitute the underlying constituents in the sentence. Morphological rules are then applied, followed by phonological or orthographic rules (for further information, see Radford, 1988, or Celce-Murcia and Larsen-Freeman, 1999).

According to Chomsky’s (1981) theory of UG, knowledge of a language consists of not only knowledge of the universal principles shared by all languages, but also knowledge of language-specific rules, or parameters of grammatical variation observed between languages or different varieties of the same language. These parameters are triggered by exposure to the target language. More recently, Chomsky (1995) has argued that ‘grammars should be described in terms of the minimal set of theoretical and descriptive apparatus necessary’ to describe a descriptively adequate depiction of linguistic phenomena (Radford, 1997, p. 265). This minimalist program of linguistic theory stems from a desire to minimize the acquisitional burden for children learning a language in a relatively brief period of time (Radford, 1997). Finally, Chomsky’s linguistic program has evolved considerably over the years. The details of this system are complex, and beyond the purview of this book. For our purposes, I will refer to this work simply as UG.

Although UG has deepened our understanding of syntax, it has been criticized for failing to account for meaning or language use in social contexts (Hymes, 1971; Halliday, 1994). In other words, UG’s focus on syntax downplayed to some extent the role of semantics, or the study of the conventional meanings of words, phrases and sentences, and excluded pragmatics, or meanings derived from context-specific use. Nonetheless, both semantics and pragmatics, together with phonology, morphology
and syntax, are critical for assessing the communicative success of an utterance within a given context.

To illustrate these shortcomings, consider the following two syntactically identical sentences.

(1.3) It is raining.
(1.4) It is working.

Both sentences begin with the pronoun subject *it*, followed by the third-person singular form of the auxiliary verb ‘be’, which carries tense (present). The auxiliary verb is followed by a main verb (*rain* or *work*) in the progressive form. Syntactically, these forms are identical apart from the different main verbs. However, there are obviously many differences in linguistic meaning and contextual use. For example, *it* in 1.3 functions as a ‘sentence filler’ in the subject position since *it* contains no referent. Sentence fillers are a resource available to those languages which require the expression of an explicit grammatical subject (e.g., English, French). However, *it* in 1.4 contains an implied referent – most likely some kind of mechanical device previously mentioned in the context. Although the semantic information and/or contextual inference related to *it* is essential for understanding these two utterances, UG fails to account for the referential information in this analysis. In short, as a model for communicative teaching and testing, the syntactocentric perspective has much to contribute; however, used alone, it may not be appropriate for all situations, and must, therefore, be adopted judiciously.

Another example of the theoretical limitations of applying a purely syntactocentric approach to L2 educational contexts is seen in the following two pairs of utterances.

**Context:** A French person, who speaks only French, is having a discussion with two Americans, who both speak English and French fluently. During the discussion, one American (Joe) lapses into English. The other American (Sue) says:

(1.5) Sue: Would you please speak French? [request and perhaps criticism]
(1.6) Joe: Oh, no problem. [acknowledgment and agreement to comply]

Later, noticing that Joe has not stopped speaking English, Sue repeats:
Sentences 1.5 and 1.7 are identical in syntactic structure. However, these sentences convey very different pragmatic information. Sentence 1.5 was intended as a polite request. It may also have encoded a hint of criticism in this context. Joe’s response in 1.6 showed that he interpreted the question as a request, and if the criticism was understood, it was ignored. However, sentence 1.7 was an exact repetition of the initial request (most likely with a moderation in prosody), stated as if Joe had not heard (or heeded) the initial request. This time, the utterance was intended as a criticism or even a chastisement for not speaking French. In 1.8 Joe responded with an apology followed by an excuse, thereby acknowledging that he had heard the sentence in 1.7 as a criticism.

In order for Joe to respond accurately, meaningfully and appropriately in each exchange, Joe had to understand the grammatical form (yes/no question) of the utterance; he had to understand the literal meaning of the words in syntax and their intended meaning in context (Speak French!), and he had to understand any additional pragmatic meanings being communicated without actually being said (criticism, annoyance). An analysis of the syntax alone would not have been able to account for the differences in the two sentences with regard to meaning.

To highlight further a need to account for meaning on a lexicogrammatical level, consider the different interpretations of the modal auxiliary ‘can’ in the following sentences:

- **Can you speak Kurdish?** (ability or potential)
- **Can I have some milk, please?** (request)
- **Can I go to the movies tonight, please?** (request for permission)
- **Can I buy you a beer?** (offer)
- **Can we talk at 10?** (suggestion)
- **Can they still be at work?** (speculation)
- **Can it get any warmer?** (theoretical possibility)

As can be seen, the modal auxiliary *can* has the same basic syntax in these instances, but the semantic representation changes. If further contextual information were provided, a host of pragmatic interpretations could be also derived from these utterances. For example, ‘Can you speak Kurdish?’ could be used (or interpreted) as a way of discrediting someone
who thinks she knows a lot about the Kurdish people. In this context, a
layer of pragmatic meaning involving negative affect and attitudinal
stance could be superimposed on the linguistic forms and literal mean-
ings. In other words, a focus on grammatical form alone may not be
efficient in L2 educational contexts to determine if L2 learners have suffi-
ciently acquired a structure to communicate effectively.

UG, like other syntactocentric theories of language, has contributed
enormously to how language teachers and testers understand language
and linguistic forms. Many L2 educators continue to draw on syntacto-
centric theories of language to design language syllabi and teach gram-
matical forms. In the same way, many language testers have designed
tests of grammatical form that are firmly rooted in these theories. In fact,
many of the traditional, multiple-choice tests of grammar are heavily
influenced by a syntactocentric approach to language. While the syntac-
tocentric theories of language continue to inform our understanding of
language structures and the principles underlying them, these theories
have fallen short on issues of meaningfulness, appropriateness, accept-
ability and naturalness – for that we might turn to corpus linguistics and
to functional grammar.

Form- and use-based perspectives of language

The three theories of linguistic analysis described thus far have provided
insights to L2 educators on several grammatical forms. These insights
provide information to explain what structures are theoretically possible
in a language. Other linguistic theories, however, are better equipped to
examine how speakers and writers actually exploit linguistic forms
during language use. For example, if we wish to explain how seemingly
similar structures like I like to read and I like reading connote different
meanings, we might turn to those theories that study grammatical form
and use interfaces. This would address questions such as: Why does a lan-
guage need two or more structures that are similar in meaning? Are
similar forms used to convey different specialized meanings? To what
degree are similar forms a function of written versus spoken language, or
to what degree are these forms characteristic of a particular social group
or a specific situation? It is important for us to discuss these questions
briefly if we ultimately wish to test grammatical forms along with their
meanings and uses in context.

One approach to linguistic analysis that has contributed greatly to our
understanding of the grammatical forms found in language use, as well as the contextual factors that influence the variability of these forms, is corpus linguistics. I will briefly describe corpus linguistics along with how findings from this approach can be useful for assessing grammar.

The common practice of compiling linguistic corpora, or large and principled collections of natural spoken and written texts, in order to analyze by computer patterns of language use in large databases of authentic texts has led to a relatively new field known as ‘corpus linguistics’. Not a theory of language per se, corpus linguistics embodies a suite of tools and methods designed to provide a source of evidence so that linguistic data can be analyzed distributionally – that is, to show how often and where a linguistic form occurs in spoken or written text. According to Biber, Conrad and Reppen (1998), these analyses typically focus on two concerns. One type of study examines the use of one linguistic feature (i.e., a lexical item or grammatical structure) in comparison with another. For example, corpus-based studies might examine the different uses of would. These studies might also compare the word wish with that-clauses and to-infinitives, or they might examine a linguistic feature with a non-linguistic feature, such as gender, dialect or setting.

Katz and Fodor (1963) looked at the connections between lexical forms and grammatical forms by examining the features of words that encode grammar. They found that in addition to encoding semantic features and restrictions, a word also contains a number of syntactic features including the part of speech (noun, verb, adjective), countability (singular, plural), gender (masculine, feminine), and it can mark prepositional co-occurrence restrictions such as when the word think is followed by a preposition (about, of, over) or is followed by a that-clause. Katz and Fodor called this ‘the grammatical dimension of lexis’. I will refer to this as lexical form, as opposed to lexical meaning.

Biber et al. (1998) identified a second kind of corpus-based study that relates grammatical forms to different types of texts. For example, how do academic texts differ from informal conversations in terms of the passive voice? Besides showing which linguistic features are possible in texts, corpus linguistics strives to identify which are probable. In other words, to what degree are linguistic features likely to occur in certain texts and in what circumstances? For example, in physical descriptions of objects the majority of the verbs are non-progressive or stative. Unlike descriptive linguistics or UG, corpus linguistics is not primarily concerned with syntax; rather, it focuses on how words co-occur with other words in a single sentence or text. In this respect, the findings from corpus-based
Based on large amounts of data, corpus linguists have begun to supply information on patterns of variation in language use, language change, and varieties of language. One type of information relates to the frequency and distribution of grammatical forms. For example, Grabowski and Mindt's (1995) study of 4,240 regular verb types found in the Brown Corpus (Francis and Kučera, 1964) and Lancaster-Oslo-Bergen Corpus (Johansson et al., 1978) discovered that regular verbs accounted for only 42.3% of the total English verb tokens, with irregular verbs making up the rest. Moreover, of these irregular verbs, 60% were accounted for by be, have, or do, and 23.6% by say, make, go, take, come, see, know, get, give, find, think, tell, become, show, leave, feel, put. In sum, these 20 verbs constituted an amazing 83.6% of the irregular verbs in the corpora. In testing grammar at different proficiency levels, this information can be very useful in helping to select appropriate content.

Besides frequency of occurrence and the distribution of forms, corpus linguistics has provided information on the different semantic functions of lexical items. For example, a corpus linguist could examine the distribution and frequency of occurrence of the word black and discover that it relates to color, race, profit, cleanliness, amount of light and so forth. Besides lexical items, corpus linguistics provides distributional and frequency information on the lexico-grammatical features of the language or those features that could be taken as both lexical and grammatical. For example, the word since has a lexical dimension given its semantic encodings and a grammatical dimension given its role as a clause marker or a preposition. This corpus-based information is of great interest to language educators because information on the distribution and frequency of grammar points helps provide an empirical basis for determining which learning points to teach or to test. (See Biber et al., 2004, for a detailed description of the spoken and written languages used at American Universities.)

With new perspectives on language use, corpus linguistics has begun to challenge language teachers to rethink how they view the content of a language curriculum and the manner in which this curriculum is presented to students. For example, instead of asking beginning and intermediate students to learn a large number of tenses and verb forms, as is done in a structural syllabus, language teachers might promote L2 vocabulary development or introduce students to features of the L2 that allow them to function appropriately in social contexts (Kennedy, 1998).