LIMITATIONS OF CONTEXT-FREE GRAMMARS

1 Linguistic Motivation

As we have seen, we can use the phrase structure trees generated by a CFG as one way of making precise the idea that what linguists would consider to be two distinct linguistic expressions can correspond to the same string. However, once we look a little more closely at some of the basic facts about natural language, some inadequacies of this approach soon become evident. Probably the most glaring one is that not all natural languages (viewed as stringsets) are context-free! Informal (and incorrect, but widely accepted) arguments to this effect were given by numerous authors starting with Chomsky in the mid-1950s. In 1982, Pullum and Gazdar showed that the published arguments against the context-freeness of NL were all incorrect, being based either on flawed mathematical reasoning or faulty linguistic data. But by 1985, Shieber presented an argument, based on simple properties of the set of context-free languages well-known to formal language theorists, that Swiss German (as a stringset) cannot be context-free. So even though most current syntactic frameworks still make use of CFGs, some additional machinery must be brought to bear. However, the various frameworks differ as to what that additional machinery should be.

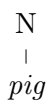
A second problem with the CFG approach to NL is that it obscures some basic questions about how the syntax of a NL relates to its phonology. In order to explain this point, we first need a bit of informal terminology. By a *(linguistic) expression*, we mean a piece of language conceived of as something that has (among other things) a semantics (the meaning that it expresses), a phonology (how it sounds), and something else that we will call its *combinatorial potential*, i.e. its ability to combine with other expressions. (Other names for this are *distribution* and *privileges of occurrence*.) The three aspects of expressions can be thought of as the linguists' counterparts, respectively, of the pronunciation, meaning, and part-of-speech information provided for words by an ordinary dictionary.

Expressions in this sense include both words (or *lexical* expressions), and more complex expressions (often called *phrases*) constructed from simpler expressions by *syntactic rules*. Thus expressions are multifaceted linguistic entities. Two other widely used terms with essentially the same sense are *sign* and *syntactic form*. The first of these, *sign*, widely used in the CG (categorial grammar) and HPSG (head-driven phrase structure grammar) commuities, is an extended usage based on Saussure's notion of a sign as a

mentally represented associative bond (French *lien de l'association*), shared among the members of a speech community, between a *signifier* (French *signifiant*, also called a *sound image*) and a signified (French *signifié*, also called a *concept*). We will avoid the term *sign* here since Saussure seems to have intended it to refer only to words (lexical expressions). The second term, *syntactic form*, is a narrowing of the more general term *(linguistic) form* used by the American structuralists (1920s through 1950s) in the sense of a piece of language with a fixed sound, meaning, and distribution. Forms in this sense include not just syntactic (or *free*) forms, but also *bound* forms, such as parts of words (both bound roots and affixes) that cannot occur on their own. We will avoid the term *form* in this sense also, because of the other widespread use of the term *form* with the sense of a superficial sensory manifestation or percept (as opposed to meaning or content or expressed concept). To summarize: in our terminology, expressions (words and phrases) are what the syntax is about, but any expression *has* a phonology and a semantics, as well as an ability to combine with other expressions to form more complex ones.

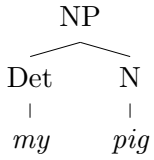
In contradistinction to syntax, the *phonology* of an NL is the system of linguistic sounds, including the ones that can be the phonologies of expressions. Just as the syntax governs the combination of words into more complex expressions, the phonology governs the combination of the most basic linguistic sound units into more complex phonological entities. The most basic sound units include not just sound segments (phonemes) that can be strung together, but also minimal prosodic units (such as pitch accents and boundary tones) from which linguistic tunes are composed, and complex phonological entities at minimum have to encompass both strings of phonemes (or maybe strings of strings of phonemes, depending on the framework) *and* tunes, as well as a specification of how the two are aligned.

Against this background, we can now explain the sense in which the use of phrase structure trees obscures basic questions about the connection betwen syntax and phonology. To begin with, linguists make a clear distinction between words and their phonologies. For example, there is a difference between the word (lexical expression) pig and the phoneme string /pIg/ which is (the segmental part of) the phonology of that word. This raises the following question: if we represent the noun pig by the tree



does the terminal node label *pig* refer to the word pig or to the phoneme string /pIg/? Strange as it might seem, syntacticians do not agree on this very fundamental question. On the one hand, it would appear that the terminals are supposed to refer to something phonological, since the terminal yield of a phrase structure tree is supposed to tell us the 'left-to-right order' (i.e. temporal succession) in which the words are phonologically manifested. On the other hand, given that two distinct but homophonous nouns, e.g. bank_1 'riverside' and bank_2 'financial institution' must be counted as different expressions (since they have different meanings), it appears that phrase structure tree representations can distinguish them only if we take the terminal node labels to be names of (lexical) expressions, not shorthands for phoneme strings:

So far so good. But now, how do we tell by looking at a more complex phrase structure tree what the expression that it represents is supposed to sound like? Consider, e.g.



How do we get from this representation to the (segmental) phonology /maj pIg/?[1] Well, first we have to assume that in addition to lexical entries such as $\mathrm{N} \rightarrow \mathsf{pig}$, our grammar (theory of the language) also specifies somewhere what each word sounds like, e.g. that the phonology of pig is /pIg/. But that alone is not enough; we also have to know that in order to determine

[1]Here we assume phonologies of expressions, written between /-lines, are strings of phoneme strings; in particular, we take word phonologies to be length-one strings of phoneme strings.

the phonology of a phrase, we have to first determine the phonologies of the daughters of the root and then concatenate them in the same order as the linear-precedence ordering of the corresponding daughters.

So far, our consideration of phrase structure trees has pushed us in the direction of a simple mathematical model of the syntax-phonology interface. Alas, it is is *too* simple, even if we ignore nonsegmental aspects of phonology. For one thing, what are we to say about forms (in the structuralist sense) that are bound (in the sense of forming a unit of segmental phonology) to a word without actually belonging to the word? Examples of this kind of form, sometimes called *phrasal affixes* or *clitics*, include the English possessive 'marker' *-'s* (sometimes called the Saxon genitive), Japanese case 'markers' such as *-ga, -ni*, and *-o*, and counterparts in many languages of (what in English grammar are usually called) 'function words' such as determiners, prepositions, complementizers, auxiliaries, and pronouns. Consider for example the sentence *Bill's pig died*. At first blush it appears that *Bill's*, pronounced /bIlz/, is a word, the 'genitive form' of the word *Bill*. This hypothesis would appear to be supported by examples like *Pete's pig died*, where the phonology of *Pete's* is /pits/: evidently the phonetic realization of the possessive marker is conditioned by the final segment of the 'stem' to which it attaches, just as if it were an inflectional suffix (such as the regular plural suffix *-s* for nouns or the regular third person singular present form *-s* for verbs). In other words, it looks as though we should account for the behavior of *-'s* not in the syntax but in the morphology (the system by which words are built up from smaller meaningful units called *morphemes*.) But then what about examples like *the mayor of Boston's hat* or *the man I was talking to's pig*? Clearly, there is something amiss with saying that *Boston's* in this example is the 'genitive form of *Boston*', or that *to's* is the 'genitive form of *to*'! Rather, it seems that what we would really like to say is that there is a *syntactic* rule that attaches *-'s* to an NP to form an expression that has the same combinatorial potential as a determiner (such as *the* or *my*). But then it seems that the *way* that *-s* combines with a NP to form a Det is different from the way that (say) a VP combines with an NP to form an S: we have to know that its phonology forms a segmental unit (what we might call a *phonological word*) together with the phonology of the last (syntactic) word in the NP.

Other examples of how phrase structure trees might be seen to provide too simple a model of the syntax-phonology interface come are provided by so-called *nonconcatenative* syntactic operations. Anyone who has studied morphology is familar with cases where certain inflected forms of words are formed not by concatenation of a prefix or suffix to a stem, but rather by

internal modification of the stem (e.g. ablaut, or shifting stress to a different syllable, or changing the tone on a syllable), by infixation (attachment of a morpheme at a phonologically defined location inside the stem), by circumfixation (wrapping of a 'discontinuous morpheme' around the stem), or reduplication (concatenation to the stem of a partial or total 'copy' of the stem). But many of these so-called morphological processes have syntactic counterparts, or at least there are phenomena which lend themselves to being analyzed in terms of such counterparts. To take just one example, one way of forming a polar (yes-no) question in Chinese can be described informally as follows: to turn a declarative sentence made up of an NP and a VP into a polar question, say the NP, followed by the VP, followed by the word *bu* 'not', followed by a copy of the head verb of the VP:

a. Ta xihuan zhei-ben shu. (lit. s/he like this-volume book)

b. Ta xihuan zhei-ben shu bu xihuan? (lit. s/he like this-volume book not like)

The point here is not to advocate a particular analysis of Chinese reduplicative questions, but just that we should not rule out, out of hand, analyses involving modes of syntactic combination whose phonological effect is something other than concatenation of the phonologies of the immediate constituents.

Examples of a different kind are provided by languages such as Swiss German and Czech which lend themselves to analyses in which some VPs are 'discontinuous' in the sense that material from 'higher' VPs can intervene between the verb and its object. Consider the following example of an English subordinate clause together with its translation equivalent in Swiss German (from Shieber 1985):

a. (that) we [let the children [help Hans [paint the house]]]

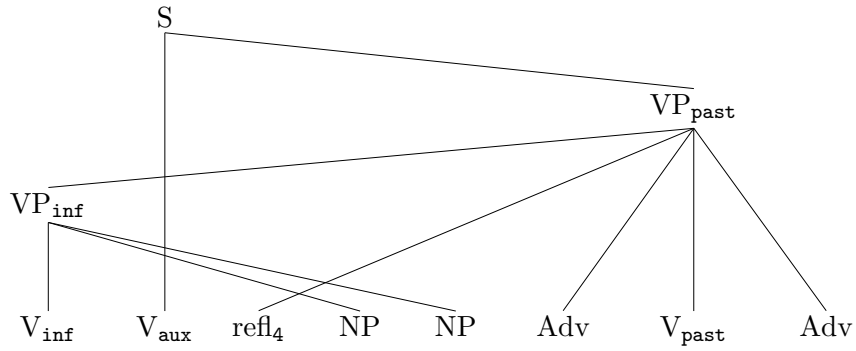
b. mer d'chind em Hans es huus lönd hölfe aastriiche

(lit. we the-children the Hans the house let help paint)

In the English sentence, what are usually analyzed as successively embedded VPs are indicated by nested pairs of square brackets. But in the Swiss German counterpart of the English VP *paint the house*, for instance, the verb *aastriiche* 'paint' is separated from its object *es huus* 'the house' by the verbs of the 'higher' VPs. Of course there are analyses (such as the LFG analysis of Bresnan et al. 1982 for similar constructions in Dutch)

which avoid such 'discontinuous constituents' by treating the grammatical relatedness of verbs and objects at a different level (viz. f-structure) from the phrase structure (LFG's c-structure).[2] But we should at least allow for the possibility of analyzing such as cases as having ordinary VP constituents whose phonological realizations are discontinuous.

Our last example of the inadequacy of traditional context-free phrase structure analysis of syntax is drawn from Czech. Speaking very informally, in this language, although the order of words within a VP is quite free, it is not usual (as it is in Swiss German) for the phonological realization of a VP to be broken up by material from outside the VP. However, there is a systematic exception to this generalization: clitics, no matter whether they are main-clause auxiliaries or object pronouns, can pile up in the so-called Wackernagel position (roughly, after the first word or after the first 'major constituent'), no matter which VPs they break up, as in the following example (Hana 2006):



Opravit	*jsem*	*se*	*mu*	*to*	včera	snažil	marně.
repair	aux$_{1sg}$	refl$_4$	to-him$_3$	it$_4$	yesterday	tried	fruitlessly

'I tried to repair it for him yesterday without success.'

Such clitic-induced multiple discontinuities have no effect on the perceived acceptability of sentences.[3]

[2] In Bresnan et al.'s analysis, the cluster of verbs form a constituent.

[3] By comparison, in Czech sentences without clitics, a single discontinuity (usually resulting from placement of 'contrastively topicalized' material at the left periphery of a VP) is allowed, but multiple discontinuities are very rare and lead to a marked decline in acceptability.