

1 Parse trees

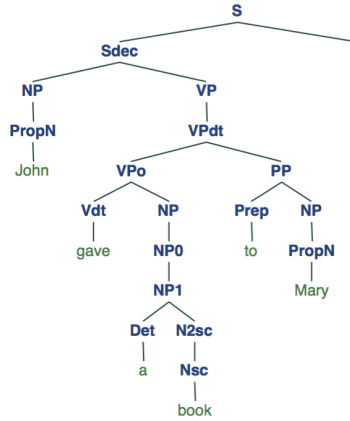


Figure 1: *John gave a book to Mary.*

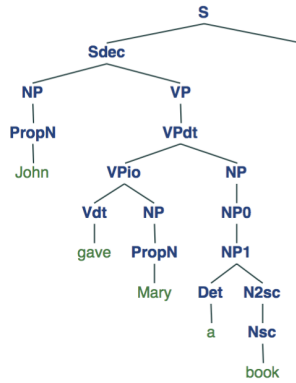


Figure 2: *John gave Mary a book.*

2 Remarks on the grammar

Design of the grammar rules

Positive aspects

Negative aspects a. **Redundancy of rules in the verbal domain**

There is a redundancy in the treatment of ditransitive verb phrases.

VPo and $VPio$ are two nodes which expand in the exact same way:

$$VPo \rightarrow Vdt NP \text{ and } VPio \rightarrow Vdt NP.$$

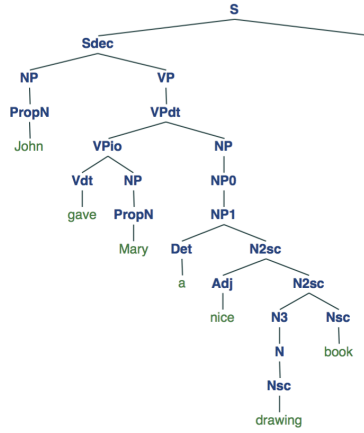


Figure 3: *John gave Mary a nice drawing book.*

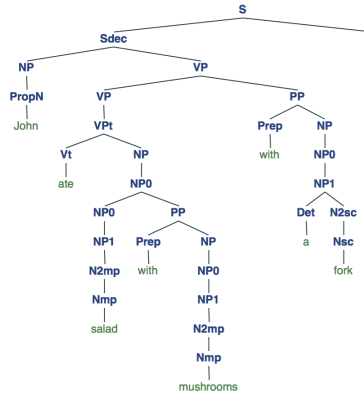


Figure 4: *John ate salad with mushrooms with a fork.*

and therefore there is no difference between them within the grammar. From a linguistic point of view this separation might have a motivation. The ditransitive verb and its direct object are represented by Po , while $VPio$ is a unit consisting of a ditransitive verb and its indirect object. As valid linguistic distinction as it might be, it does not need to be included in simple grammar like the one discussed here. In particular, no use is made of the distinction between that would limit overgeneration. If that was the case, then having both non-terminals would be a valid design choice. For instance, one might consider putting some restrictions on what kinds of NP can be direct or indirect objects. The needless distinction between VPo and $VPio$ propagates up to the topmost VP expansion rules. As a result, there are more VP-related rules than necessary. The grammar

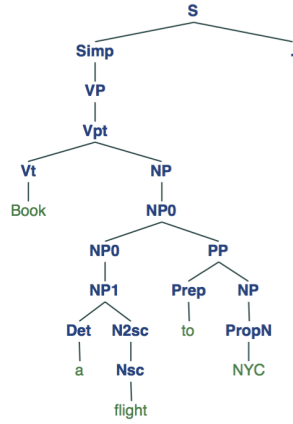


Figure 5: *Book a flight to NYC.*

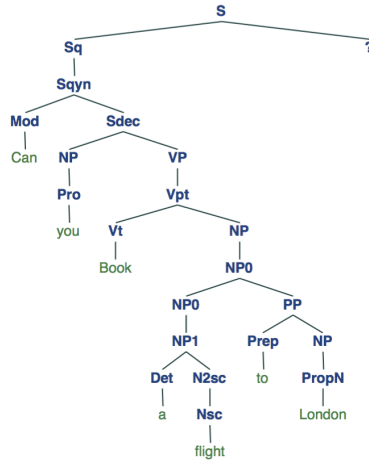


Figure 6: *Can you book a flight to London?*

contains

$$\begin{aligned}
 VP &\rightarrow VP_i \mid VP_t \mid VP_{dt} \mid Mod \ VP \mid VP \ Adv \mid VP \ PP \\
 VP_{dt} &\rightarrow VP_o \ PP \\
 VP_{dt} &\rightarrow VP_{io} \ NP \\
 VP_o &\rightarrow V_{dt} \ NP \\
 VP_{io} &\rightarrow V_{dt} \ NP
 \end{aligned}$$

but it could have contained as little as three rules instead:

$$\begin{aligned}
 VP &\rightarrow VP_i \mid VP_t \mid VP_{dt} \mid Mod \ VP \mid VP \ Adv \mid VP \ PP \\
 VP_{dt} &\rightarrow VP_d \ NP \mid VP_d \ PP
 \end{aligned}$$

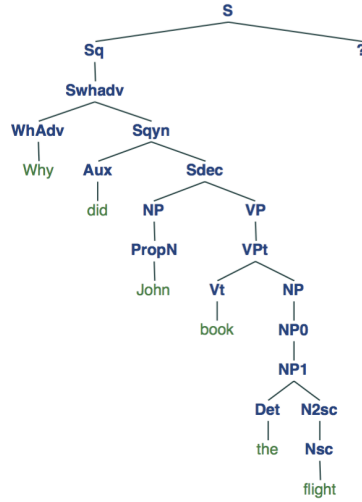


Figure 7: *Why did John book a the flight?*

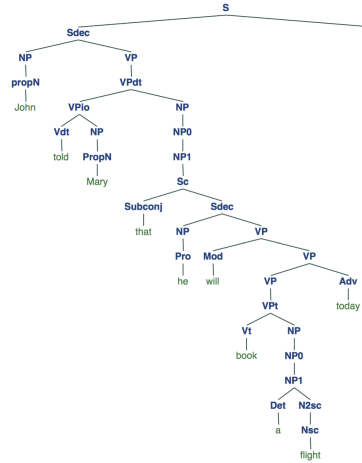


Figure 8: *John told Mary that he will book a flight today.*

$$VPd \rightarrow Vdt NP$$

b. **Redundancy of rules in the nominal domain**

Similarly, the NP0 non-terminal node seems redundant. It's purpose is to allow for recursive generation of prepositional phrases following a noun phrase. This can be achieved by accounting for this kind of recursion within one of the existinf NP rules. We could propose:

$$\begin{aligned} NP &\rightarrow \text{PropN} \mid \text{Pro} \mid \text{NP1} \mid \text{NP PP} \\ \text{NP1} &\rightarrow \text{Det N2sc} \mid \text{N2mp} \mid \text{Sc} \mid \text{NP1 PP} \end{aligned}$$

The first option has the advantage of allowing for proper names and pronouns to be modified by prepositional phrases, which could be desirable in a more comprehensive grammar, however is not necessary for the grammar at hand. The second proposition is more conservative in that it would not lead to generating any string the original grammar doesn't. The

$$NP0 \rightarrow NP1 \mid NP0 PP$$

rule expands to an *NP1* followed by an arbitrary number of *PPs*. The same effect will be achieved by adding the *NP1* \rightarrow *NP1 PP* expansion, with the benefit of reducing the number of non-terminals. One might argue that the *NP0* node corresponds approximately to an *X'* node in *X-bar* syntax, and hence is theoretically justified. This, however, would need to mean that *PPs* to the right of the head noun are adjuncts rather than complements of that noun, which is not true. [TODO citation]

c. **Doubtful usefulness of the *Sq* node**

The *Sq* node serves no purpose other than to reflect the conceptual relation between *Sqyn* and *Swadv*, namely the fact both are non-terminals expanding into questions. This, however, represents only a superficial, if not mistaken, grammatical insight. In fact *wh*-questions and auxiliary-inversion questions are represented by very different structures (*WH*-phrases and *Aux*-phrases respectively[TODO citation]). If the motivation for including the *Sq* node is linguistic, then the reasons seem insufficient. If, on the other hand, the node was included in the grammar to simplify the expansion of *S*, then by analogy we should also include

$$Sd \rightarrow Sdec \mid Simp$$

and change the expansion of *S* to

$$S \rightarrow Sd \text{ '}' \mid Sq \text{ '}'$$

d. **Lower case and capitalized versions of terminals.**

It does not seem necessary to include lower case and capitalized versions of the same terminals. In the sample sentences we have seen *Book*, *Can*, and *Why* being capitalized to account for their sentence-initial positions. Instead of duplicating lexical items, we could simply capitalize the first word of a generated sentence and ignore sentence-initial capitalization during parsing. There would be no ill effects of such a change, since every generated sentence is, by definition, grammatical. Every word which happens to be initial in a sentence produced by the grammar is allowed to be there. The additional confirmation of this fact in form of the word's capitalised version being present in the grammar is unnecessary.

Overgeneration a. Lack of subject-verb agreement

There's distinction between single/count and plural/mass nouns, but no corresponding distinction in the verbal domain. The problem is somewhat masked by the fact that most of the included verbs are in past tense, where in English there are no subject-verb agreement phenomena. Nevertheless, we can see that the grammar would accept **the fork book a flight.*, a sentence which ignores verb conjugation requirements.

b. Treating VP as full sentences

The rules:

$$S \rightarrow \text{Simp} \text{ '.'}$$

$$\text{Simp} \rightarrow \text{VP}$$

imply that any *VP* can be a grammatical sentence on its own. This, however, is true only of *VP* in imperative mood. The grammar, however, allows for treating any *VP*, e.g. **ate today*, as a full sentence. Solving this problem is not possible without substantially expanding the grammar. We would have to distinguish between infinitival and inflected forms of verbs, and extend the set of possible expansions of *VP* accordingly. We could then state that *Simp* subcategories for *VPs* whose head is an infinitival form of a verb. Therefore, we would need the following rules:

$$Vt_{inf} \rightarrow \text{'Book'} \mid \text{'Tell'} \mid \dots$$

$$VPi_{inf} \rightarrow \text{'Eat'} \mid \dots$$

$$VPt_{inf} \rightarrow Vt_{inf} \text{ NP} \mid VPt_{inf} \text{ Adv} \mid VPt_{inf} \text{ PP}$$

$$\text{VP} \rightarrow VPt_{inf}$$

$$\text{Simp} \rightarrow VPt_{inf} \mid VPi_{inf}$$

Even then, additional restrictions would be needed to account for characteristics of the imperative mood, e.g. the fact that the second person pronoun needs to be in its reflexive form when used as a direct object of the matrix verb [TODO citation]. Otherwise the grammar would generate sentences such as **Book you a flight to London.* instead of grammatical **Book yourself a flight to London.*

c. Case agreement

The grammar does not account for case subcategorisation requirements of verbs. The only corner of the modern English syntax which still exhibits case phenomena is the pronoun system[TODO citation], however the grammar does not include a *ProNom* and a *ProAcc* preterminals. Having no distinction between nominative and accusative forms leads to admitting non-grammatical sentences, e.g. **Mary gave he a nice salad.* instead of *Mary gave him a nice salad.*

d. Verb and noun subcategorization for pronouns

Verbs and nouns tend to co-occur with particular pronouns more often than with others. However, the grammar does not account for this preference. For instance, one can *eat with a fork*, *eat with Mary*, or have a *flight to London*, but the grammar also allows for having a **salad to fork* and booking a **flight with NYC*.

e. **Multiple modals**

The grammar allows for an unlimited number of modals to precede a VP. This can lead to generating sentences such as **Can will can can John book a flight?*. However, English limits the acceptable number (up to 4, depending on dialect), orderings (evidential, epistemic, deontic), and identity (*might* or *may* being the most commonly first in the sequence (Di Paolo 1989)) of verbs in multiple modal constructions. In order to capture these restrictions we would need a whole series of serially connected expansion rules, each introducing one modal of a certain type.

3 Ambiguity Analysis

3.1 "John gave a book to Mary."

Distinct Parse(s): 3

Ambiguity Source:

3.2 "John gave Mary a book."

Distinct Parse(s): 1

Ambiguity Source:

3.3 "John gave Mary a nice drawing book."

Distinct Parse(s): 2

Ambiguity Source:

3.4 "John ate salad with mushrooms with a fork."

Distinct Parse(s): 5

Ambiguity Source:

3.5 "Book a flight to NYC."

Distinct Parse(s): 2

Ambiguity Source:

3.6 "Can you book a flight to London?"

Distinct Parse(s): 2

Ambiguity Source:

3.7 "Why did John book the flight?"

Distinct Parse(s): 1

Ambiguity Source:

3.8 "John told Mary that he will book a flight today."

Distinct Parse(s): 3

Ambiguity Source: