

# 1 Project Demonstration

## 1.1 Instances Correctly Accepted and Rejected

Examples of accepted and rejected sentences for all categories listed in the handout are shown in Table 1, and tests are carried out using our own CKY algorithm. We have excluded accents in my grammar.

Accepted/Rejected	Sentence	Category
Accepted	je regarde la television	Basic Sentence Order
Rejected	je la television regarde	Basic Sentence Order
Accepted	je regarde la television	Subject-verb Agreement
Rejected	je regardes la television	Subject-verb Agreement
Accepted	le chat ne mange pas le poisson	Negation
Rejected	le chat mange ne pas le poisson	Negation
Accepted	je regarde Jonathan	Proper Names without the Determiner
Rejected	je regarde le Jonathan	Proper Names without the Determiner
Accepted	je regarde le Canada	Proper Names with the Determiner
Rejected	je regarde Canada	Proper Names with the Determiner
Rejected	je regarde la chat	Determiner-noun Agreement
Accepted	il la regarde	Direct object pronouns
Rejected	il regarde la	Direct object pronouns
Accepted	je regarde le chat noir	Adjective Position
Rejected	je regarde le noir chat	Adjective Position
Accepted	je regarde le beau chat	Adjective Position
Rejected	je regarde le chat beau	Adjective Position
Accepted	je regarde la derniere semaine	Adjective Position
Accepted	je regarde la semaine derniere	Adjective Position

Table 1: Examples of Accepted and Rejected Sentences

## 1.2 Instances of Overgeneration and Undergeneration

One possible cause of overgeneration is to have intransitive verbs. In our grammar, we have contained one intransitive verb, "vais," which means "go." Since the grammar does not distinguish between transitive and intransitive verbs and does not contain prepositions, it is possible to generate "je vais le Canada," which is ungrammatical. The proper way to say this is "je vais au le Canada."

The CFG we have only covers a small portion of the French grammar. For example, we cannot model tenses, and sentences such as "je la a mangée" would be regarded as ungrammatical. In addition, the grammar cannot accommodate multiple adjectives. For instance, "le beau chat noir," which translates to "the beautiful black cat," is not supported by our grammar.

# 2 Discussions

## 2.1 Advantages of Modeling French with CFG

As to the CFG per se, one advantage is that the grammar rules are easy to write. Combined with the facts that French is a well-defined language with few grammatical exceptions, we can cooperate with linguists to produce rigid grammar rules. Notice that for analytic languages such as Chinese, the grammar is highly flexible and thus hard to define. Also, with the CKY algorithm, we can parse a constituent tree from a sentence automatically.

## 2.2 Disadvantages of Modeling French with CFG

One disadvantage of modeling French with CFG is the enormous number of non-terminals. French is highly-inflected. To incorporate noun declensions and verb conjugations, we have to make lots of distinctions among nouns and verbs, resulting in more symbols and rules. For instance, to model nouns' genders, we have to rewrite each related rule with gender tags added. Likewise, distinguishing verb conjugations lets our rule sets grow exponentially. In addition, CFG itself is not powerful enough. For instance, CFG does not consider probabilities of each parsed tree, and thus it is impossible for us to pick

the most likely parsed tree; this problem can be resolved via PCFG. Another issue with CFG is that it is cumbersome to model more flexible grammar rules. If a grammar rule requires the knowledge of its surrounding contexts, we can only accommodate this via annotations, which will inevitably create more rules.

### **2.3 Some Aspects of French Grammar not Captured by my CFG**

As mentioned before, our CFG can only model a small subset of the French grammar. There are many unsupported features, such as tenses, interrogative and imperative sentences, subordinate clauses, etc. Concretely, as mentioned before, my CFG does not allow multiple adjectives in a single noun phrase. Any sentence consisting of "le beau chat noir" shall be rejected. Also, word tenses are not accommodated. Simple sentences like "je la a mangée" shall be deemed ungrammatical. In addition, due to the lack of propositions, our CFG cannot model intransitive verbs correctly.