

### Prerequisites

- general(abbreviations[w.l.o.g.])
- sets(notation, operations)
- strings(notation)

## Proof: Equivalence of positive and negative grammars

This section defines both negative and positive versions of  $n$ -gram grammars and shows that they are expressively equivalent. Like in the section on the equivalence of mixed and fixed  $n$ -gram grammars, this is accomplished by a **constructive** proof. A proof is constructive if it doesn't just derive the existence of some object, but gives a concrete procedure for constructing this object. In the case at hand, the proof shows how to construct a positive grammar from a negative one, and the other way around.

---

**Definition 1.** Let  $\Sigma$  be some alphabet, and  $\Sigma_E$  its extension with edge marker symbols  $\bowtie, \bowtie \notin \Sigma$ . An  $n$ -gram over  $\Sigma_E$  is an element of  $\Sigma_E^n$  ( $n \geq 1$ ). An  $n$ -gram grammar  $G$  over alphabet  $\Sigma$  is a finite set of  $n$ -grams over  $\Sigma_E$ . Every  $n$ -gram grammar has a **polarity**:

- If  $G$  is negative (also denoted  $^-G$ ), then a string  $s$  over  $\Sigma$  is well-formed with respect to  $^-G$  iff there are no  $u, v$  over  $\Sigma_E$  and no  $g \in ^-G$  such that  $\bowtie^{n-1} \cdot s \cdot \bowtie^{n-1} = u \cdot g \cdot v$ .
- If  $G$  is positive (also denoted  $^+G$ ), then a string  $s$  over  $\Sigma$  is well-formed with respect to  $^+G$  iff for all  $u, v$  over  $\Sigma_E$  and  $g \in \Sigma_E^n$  such that  $\bowtie^{n-1} \cdot s \cdot \bowtie^{n-1} = u \cdot g \cdot v$ , it holds that  $g \in ^+G$ .

The **language of**  $G$ , denoted  $L(G)$ , contains all strings that are well-formed with respect to  $G$ , and only those.

---

**Theorem 2.** For every  $n$ -gram grammar  $G$  there exists a grammar  $G'$  of opposite polarity such that  $L(G) = L(G')$ .

*Proof.* We assume w.l.o.g. that  $G$  is a positive grammar and denote it by  $P$ . We define a negative counterpart  $N$  as  $\Sigma_E^n - P$  and show that  $L(P) = L(N)$ .

First, every  $s \in L(P)$  is necessarily a member of  $L(N)$ . Assume towards a contradiction that  $s \notin L(N)$ . Then there must be some  $g \in N$  such that  $\bowtie^{n-1} s \bowtie^{n-1} = u \cdot g \cdot v$  ( $u, v \in \Sigma_E^*$ ). But since  $N := \Sigma_E^n - P$ ,  $g \in N$  implies  $g \notin P$ , wherefore  $s \notin L(P)$ . As this contradicts our initial assumption that  $s \in L(P)$ , it cannot be the case that  $s \notin L(N)$ . So  $s \in L(N)$  after all.

In the other direction, suppose that  $s \notin L(P)$ . Then by definition there are  $u, v \in \Sigma_E^*$  and  $g \in \Sigma_E^n \ni g \notin P$  such that  $\bowtie^{n-1} s \bowtie^{n-1} = u \cdot g \cdot v$ . But then  $g \in N$ , which entails  $s \in L(N)$ .

This shows that  $s \in L(P)$  iff  $s \in L(N)$ . As  $s$  was arbitrary, this holds for all strings and establishes  $L(P) = L(N)$ , which concludes our proof.