Stop word removal ∼ phonological tiers (Solutions)

Exercise 1.
Write an 11-gram grammar that generates *hasxintilawas* and *haʃxintilawaʃ*, but not *hasxintilawaʃ* and *haʃxinitilawas*. The grammar may be positive or negative, whichever you prefer.

Solution

The exercise leaves open whether *hasxintilawas* and *haʃxintilawaʃ* should be the only two strings generated by the grammar, or simply two among many strings generated by the grammar. Let us assume the latter. Then all we need is a negative 11-gram grammar that consists of the following two 11-grams:

- *sxintilawaʃ*
- *ʃxintilawas*

Exercise 2.
Extend the grammar so that it also captures the fact that *ʃtajanowonowaʃ* is licit whereas *stajanowonowaʃ* is illicit. You might have to move beyond 11-grams.

Solution

Since negative grammars can have n-grams of mixed length, we keep the previous grammar and add two forbidden 14-grams:

- *ʃtajanowonowas*
- *stajanowonowaʃ*

Exercise 3.
Carry out the same calculations for

- *haʃxintilawaʃ*, and
- *haʃxintilawas*, and
- *ʃtajanowonowaʃ*.

Solution

Tier			ʃ										ʃ
Word	h	a	ʃ	x	i	n	t	i	l	a	w	a	ʃ

Tier			ʃ										s
			|									|
Word	h	a	ʃ	x	i	n	t	i	l	a	w	a	s

Tier	ʃ													ʃ
	|													|
Word	ʃ	t	a	j	a	n	o	w	o	n	o	w	a	ʃ

Exercise 4.
As an abstract example, suppose that our alphabet consists of a, b, and c, and that all symbols except c should be projected on the tier. What is the tier of *aabaccacb*?

Solution

The tier is *aabaab*, as illustrated below:

Tier	a	a	b	a			a		b
	|	|	|	|			|		|
Word	a	a	b	a	c	c	a	c	b

Exercise 5.
Compute the values for all of the following:

1. $del_{-\{a,b\}}(aaccbad)$
2. $del_{+\{a,b\}}(aaccbad)$
3. $del_{+\{a,b\}}(aababad)$
4. $del_{-\{a,b\}}(aababad)$
5. $del_{-\{a,b\}}(\varepsilon)$
6. $del_{+\{a,b\}}(\varepsilon)$

Solution

1. $del_{-\{a,b\}}(aaccbad) = ccd$
2. $del_{+\{a,b\}}(aaccbad) = aaba$
3. $del_{+\{a,b\}}(aababad) = aababa$
4. $del_{-\{a,b\}}(aababad) = d$
5. $del_{-\{a,b\}}(\varepsilon) = \varepsilon$
6. $del_{+\{a,b\}}(\varepsilon) = \varepsilon$

Explanation

The important thing to keep in mind is whether the set specifies what symbols to remove, or what symbols to keep. For the empty string the distinction is immaterial as the empty string contains no symbols to remove nor to keep.

Exercise 6.
Explain why this holds. Illustrate your argument with a few examples.

Solution

Given Σ and A, let $B := \Sigma - A$. Then it must hold for every string s that $del_{-A}(s) = del_{+B}(s)$ and $del_{+A}(s) = del_{-B}(s)$. This is because removing all members of A from s is the same as keeping all symbols that are not members of A, but by our definition of B that's exactly those symbols that are elements of B. Similarly for keeping members of A, which is the same as removing symbols in $\Sigma - A$, which is our set B.

As a concrete example, suppose $\Sigma := \{a, b, c\}$, we have some string $s := aabcacba$, and $A := \{a\}$. Then $B := \{b, c\}$ by our definition. And it holds both that $del_{+A}(s) = aaaa = del_{-B}(s)$ and that $del_{-A}(s) = bccb = del_{+B}(s)$.

Exercise 7.
The term **culminativity** refers to the property that every word has exactly one primary stress. Suppose that our alphabet is $\{\sigma, \acute{\sigma}\}$, where σ denotes an unstressed syllable and $\acute{\sigma}$ one with primary stress. Specify a set ^{+}T of tier symbols and a bigram grammar G to capture culminativity (*hint*: $\rtimes$ and $\ltimes$ can be used with tiers, too).

Solution

The set of tier symbols is $^{+}T := \{\acute{\sigma}\}$, and $G := \{\rtimes\ltimes, \acute{\sigma}\acute{\sigma}\}$ a negative bigram grammar.

Explanation

For culminativity, we do not care at all about unstressed syllables (σ), so it is sufficient to project only stressed syllables ($\acute{\sigma}$) onto the tier. A well-formed tier must consists of exactly one stressed syllable and nothing else. In other words, the tier must not be the empty string, which we prevent with the forbidden bigram $\rtimes\ltimes$, and the tier must not contain more than one stressed syllable, which we prevent with the forbidden bigram $\acute{\sigma}\acute{\sigma}$.

For example, the string $\sigma\sigma\sigma\sigma\acute{\sigma}\sigma\sigma$ satisfies culminativity, and its tier is $\acute{\sigma}$, which is allowed by our grammar. On the other hand, the illicit string $\sigma\sigma\sigma\sigma\sigma$ has an empty tier, which we rule out, and the illicit string $\acute{\sigma}\sigma\sigma\sigma\acute{\sigma}\sigma$ has the tier $\acute{\sigma}\acute{\sigma}$, which we do not allow either. The same is true for the illicit $\acute{\sigma}\acute{\sigma}\sigma\acute{\sigma}\acute{\sigma}$, the tier of which is $\acute{\sigma}\acute{\sigma}\acute{\sigma}\acute{\sigma}$ and thus still contains an instance of the forbidden bigram $\acute{\sigma}\acute{\sigma}$.