

Processing Silesian on the computer

Petr Homola

December 29, 2023

Summary

This document formally describes the linguistic properties of the Silesian language and illustrates with examples how natural language processing of the language can be implemented.

Contents

1	Preliminaries	1
2	Morphonemics	2
3	Syntax	2

This is a draft.

1 Preliminaries

Upper Silesian is a pluricentric West Slavonic language, whose morphology, syntax and vocabulary are transitional between Czech and Polish. For the purposes of this paper, *Silesian* refers to the group of dialects spoken in Teschen Silesia except for the markedly distinct dialects in the region's South West, which we refer to as *Moravian*.

2 Morphonemics

The basic unit that can distinguish words from one another is the **phoneme**. In Silesian, /g/, /h/ and /ch/ are all phonemes, as illustrated with the following triplets: *grómy* ‘we play’, *chrómy* ‘lame’, *hrómy* ‘thunders’; *gore* ‘burns’, *chore* ‘ill’, *hore* ‘upwards’.

A **morpheme** is the smallest component of a word that can carry meaning. For example, in the word *nimocny* ‘ill’ there are three morphemes: *ni* ‘not’, *moc* ‘power’ and *y*, a morphological marker for adjectives (masculine, nominative, singular).

A **morphoneme** is a set of phonemes that alternate in a morpheme. For example, in /*las*/ ‘forest’ (nominative) and /*leše*/ ‘forest’ (locative) there is the morphoneme {*s,ś*}.

An **allomorph** is a phonologically conditioned variant of a morpheme. For example, in the word *nimocny* above, *ni* is the allomorph of the negative prefix that appears before nasal consonants. The other allomorph, appearing in other contexts, is *nie*, as in *niewielki* ‘not big’.

3 Syntax

Silesian is a non-configurational language, which means that syntactic relations are expressed morphologically (via inflection) rather than via word order, which is fairly free at the clause level and mostly determined by topic-focus articulation.

The basic schema of a Silesian declarative sentence is as follows:¹



Figure 1: Schema of a Silesian sentence.

In the schema X is a phrase, such as a noun phrase or a verb phrase, and Aux is an enclitic auxiliary, for example:²

¹The corresponding phrase structure rule is

$$S \rightarrow X (\text{Aux}) X^*$$

²Auxiliaries can also be attached directly to a word, for example *jo zech* is equivalent

- (1) *Jo żech spał całóm noc*
 I AUX,1SG slept whole night
 "I was sleeping whole night."

The phrase structure of the sentence is as follows:

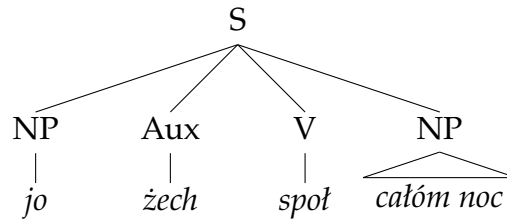


Figure 2: Parse of *Jo żech spał całóm noc*.

More interesting is the dependency tree of the sentence:

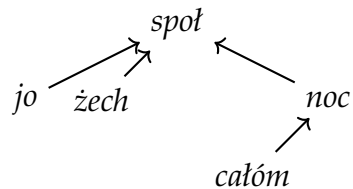


Figure 3: Dependency tree of *Jo żech spał całóm noc*.

As can be seen, the verb is the root of the sentence, on which the other phrases depend. The auxiliary does not contribute anything to the semantics of the sentence, it is merely a morphological marker.

to *joch*. Due to the principle of lexical integrity, this might cause problems in a phrase structure grammar, which we shall get back to later.