Distribution of Polish Passives in a Universal Dependency Treebank using GREW

Author: Daniel Gallagher

Date: February 24, 2025

1 Dependency Graph

A dependency graph establishes relations between the words of a sentence in a way which allows the quick dependency tagging of large corpora, and which allows easy cross-linguistic comparison through the use of *Universal Dependencies*. In this first part of the assignment, a sentence was chosen from the book *"Slaughterhouse-Five"* by Kurt Vonnegut:

(1) "Billy Pilgrim opened his eyes in the hospital in Vermont."

The dependency graph for this sentence can be viewed in Figure 1.

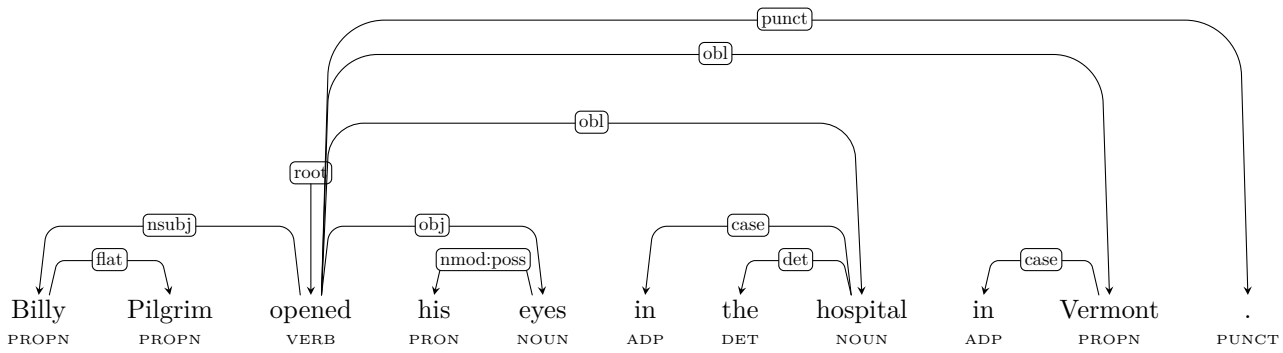

Figure 1: Dependency Graph Example

2 GrewMatch

GrewMatch is a web application which allows us to make graph requests among treebanks using the query language *grew*. In this assignment we will examine

the frequency of different passive constructions in Polish through the use of grew corpus queries. This does not serve as a thorough investigation but rather a brief analysis which shows the power of corpus querying for linguistic investigation.

2.1 Forms of Polish Passives

Verbs in Polish can have two different types of voices: *active* and *passive*. The active voice is the most common, where the subject of the sentence performs the action expressed by the verb. The passive construction is used to place the focus on the receiver of the action rather than on the subject. Passive constructions can be formed using a number of methods in Polish. These include a Verb + Reflexive, or vice versa, as well as the use of an auxiliary verb such as "to be" or "to stay" with a passive particle. Sentences (2) to (5) provide respective examples taken from (Szczepanska 2012).

(2) "Most się budował."
Bridge REFL built-IMPF
"A/The bridge was built."

(3) Dobrze się mieszka.
Well REFL lives
"It is good to live here."

(4) (Ja) jestem kochany.
(I) am loved-*PERF*
"I am loved."

(5) Szkoła została zamknięta.
school stayed-FEM closed-PERF
"The school was closed."

We will take a look at the frequency of these forms of passive constructions in Polish using grew and attempt to make some generalisations about which are most commonly used and which are not.

2.2 Constructing GREW Queries

The treebank used for this analysis was *UD_Polish-LFG@2.13*, in total consisting of 17,246 sentences. We can use GREW in order to query a dataset to investigate the frequency of each type of passive. The queries constructed for each passive, respectively, are indicated below:

1. pattern { A [upos="VERB", Person="3"]; B [form="się"]; A < B;}
2. pattern { A [form="się"]; B [upos="VERB", Person="3"]; A < B; }
3. pattern { A [lemma="być"]; B [upos="ADJ", VerbForm="Part", Voice="Pass"]; A < B; }

4. pattern { A [lemma="zostać"]; B [upos="ADJ", VerbForm="Part", Voice="Pass"]; A < B;}

In the first case, we have the Verb + REFL construction. We see that the tense is set to *present* and the person is set to *3*. This is to rule out examples where a reflexive must be used with the verb for its regular meaning. However, not all examples of this have been separated and more on this topic will be discussed in Section 2.3. Note that the reflexive marker is the same regardless of person or number. The second query is the similar to the first, though with the order of the verb and the reflexive swapped. The third searches for a passive construction using *być* (to be) and the last using the verb *zostać*. We can see that the *VerbForm* and *Voice* arguments are useful here in order to specify that we are searching for a passive participle. In each query we see the statement *A < B;*, which is simply indicating that the first relevant word being searched in our query (A) must come before the second (B).

2.3 Results

The results of the GREW queries can be seen in Table 1. In this table is represented each type of construction, along with the number of occurrences in the dataset. The total number of passive constructions found in the dataset is 895, which is 5.19% of the total sentences in the dataset. The percentage indicated in the table is the percentage of a given construction over all passive constructions found. We see that the most occurrences by far is the verb followed by the reflexive *się* at 656, or 63.94% of all passive constructions found. One reason for this is that this construction can also be used for other purposes in Polish too, for instance in the following example:

(6) "Greg myje się."
Greg washes REFL
"Greg washes himself."

However, it is also likely to be still the most common way of constructing the passive. The difficulty of separating examples like these from examples of passive constructions which use the reflexive is an issue that is out of the scope of this assignment. The reverse construction (REFL + verb) is significantly less common at 13.65% of all passive constructions (140 occurrences). This is still a significant amount and shows the relaxed worder in Polish when compared with English. The next most common type is the use of być ("to be") with the passive participle, totalling 12.87%. This is a few percentage points higher than the same construction but with the alternative verb zostać ("to stay"), the occurrences of which totalling to 9.56% of the constructions found.

2.4 Discussion

The results have shown us that the Verb + REFL construction appears to be the most commonly used way of expressing the passive in Polish. However,

Position 1	**Position 2**	**Occurrences**
Verb	Reflexive (*się*)	656 (63.94%)
Reflexive (*się*)	Verb	140 (13.65%)
Be (*być*)	Passive Participle	132 (12.87%)
Stay (*zostać*)	Passive Participle	98 (9.56%)

Table 1: Occurrences of Verb Forms

further work must be done to separate it from other uses. Another find is that the construction with the reverse order of REFL + Verb is far less common in Polish. Given Polish is a language without strict word order, it is surprising to see such a disparity. Another insight is that the auxiliary verb "be" is more commonly used to represent the passive than "zostać". This may be because the latter is more likely to be used only if there is a change in state whereas the former can be more generally used.

References

Szczepanska, Natalia (2012). "The passive construction być/zostać + V-n-/-on-/-t- in Polish : an aspectual analysis". PhD Thesis. University of St Andrews.