

Discontinuous noun phrases containing adjectival or adjective-like modifiers in Middle Polish texts.

Preliminary research conducted on an experimental
dependency treebank

Aleksandra Wieczorek

Adjectival and adjective-like modifiers (AM)

- adjectives

- ***ciepłego*** *chleba*
warm.ADJ.GEN.SG.M bread.N.GEN.SG
'of warm bread'

- adjective numerals

- ***drugiego*** *listy*
second.ADJ.NOM.SG.M letter.N.NOM.SG
'second letter'

- adjective pronouns

- ***tego*** *dnia*
this.ADJ.NOM.SG.M day.N.NOM.SG
'on this day'

Adjectival and adjective-like modifiers (AM)

- passive adjective participles

- *z szablą* ***dobytą***
with saber.N.INS.SG bring out.PPART.INS.SG.F
'with a brought out saber'

- active adjective participles

- *nieprzyjacielowi* ***atakującemu***
enemy.N.DAT.SG attack.APART.DAT.SG.M
'to the attacking enemy'

- past adjective participles

- *zębu* ***spróchniałego***
tooth.N.GEN.SG decay.PASTP.GEN.SG.M
'of a decayed tooth'

Adjectival and adjective-like modifiers (AM)

- abbreviations

- *ś* *Franciszka*

- S. (Saint) Francesco.N.GEN.SG

- ‘of saint Francesco’

- digits

- *rozdziału* **3**

- chapter.N.GEN.SG 3.

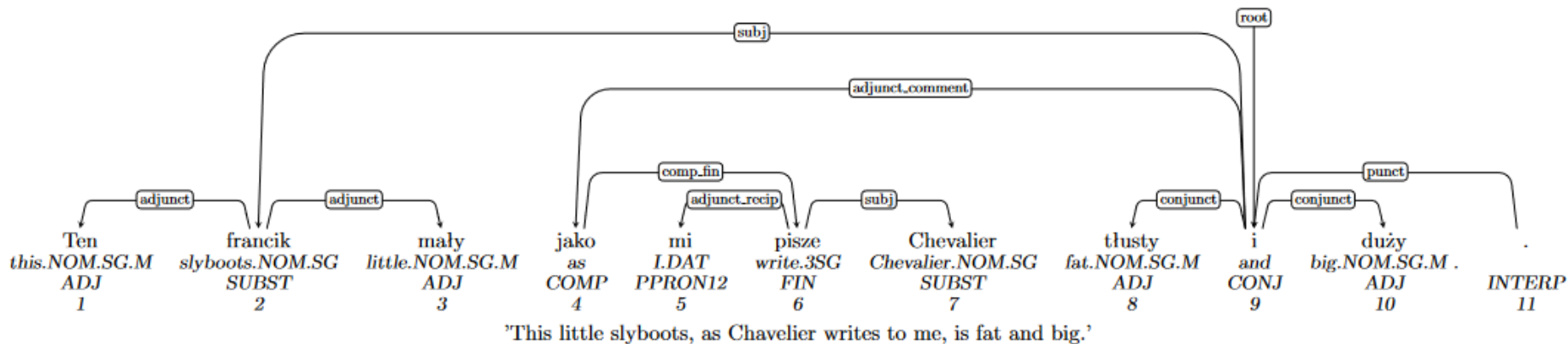
- ‘of the third chapter’

- noun phrases in genitive case
 - *syn* ***jej***
 son.N.NOM.SG she.PPRON3.GEN.SG
 ‘her son’
 - ***czytelników*** *ciekawość*
 reader.N.GEN.PL curiosity.N.NOM.SG
 ‘readers’ curiosity’
- prepositional phrases
 - *prawo* ***o*** ***polowaniu***
 law.N.NOM.SG about hunting.N.LOC.SG
 ‘hunting law’

Discontinuous noun phrases containing AM

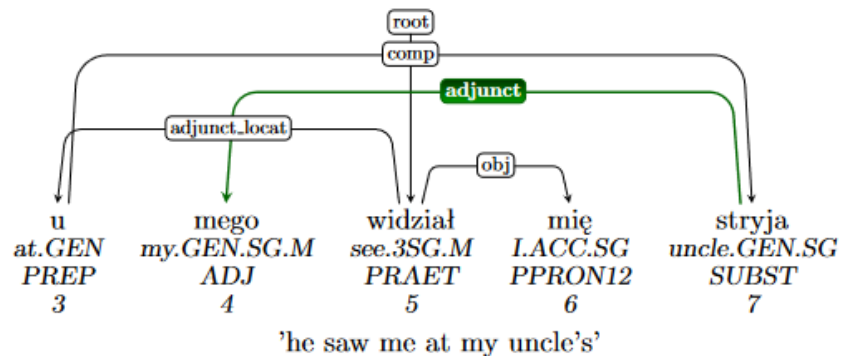
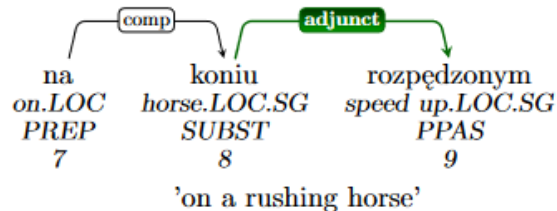
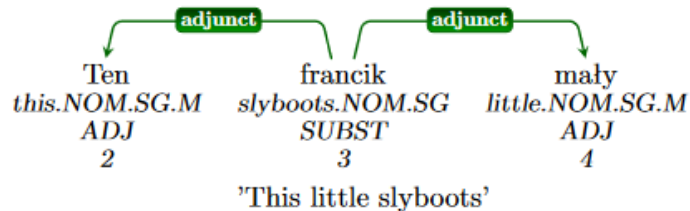
- Noun and AM are separated by another sentence component:
 - *u mego* *widział* *mię* *stryia*
at **my.GEN.SG.M** see.PRAET.3SG.M I.ACC.SG **uncle.GEN.SG**
'he saw me at **my uncle's**'
- More frequent in 17th and 18th centuries than today (e.g. Karplukówna 1967; Perczyńska 1978; Wiśniewska 1994; Wyderka 2005; Śliwiński 2006).

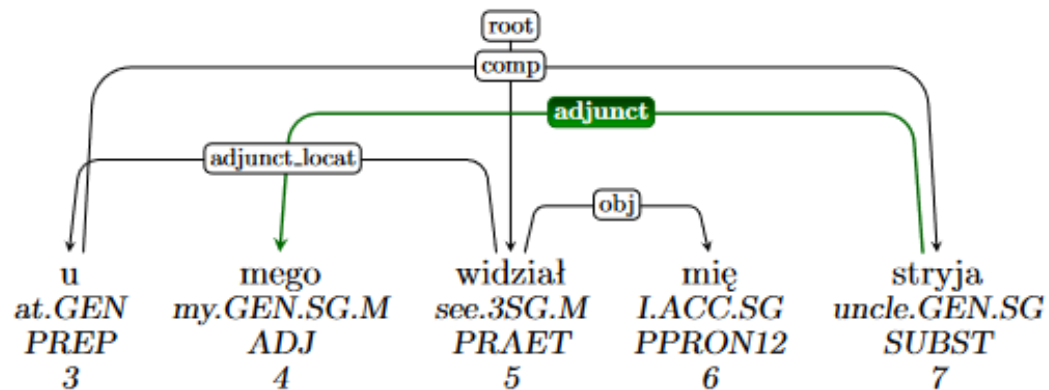
Data set: a syntactically annotated corpus



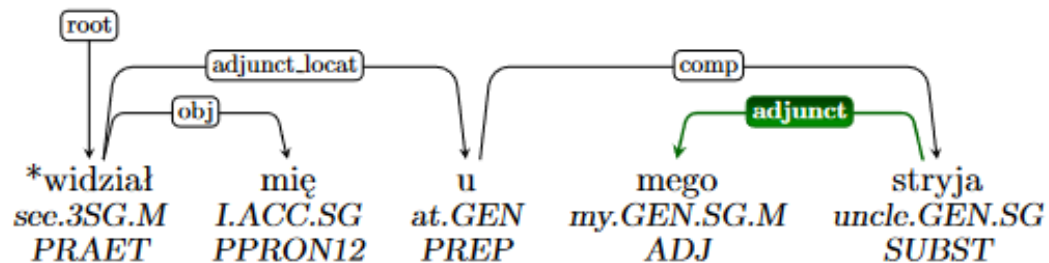
Jan III Sobieski, *Letters to Marysieńka*, 1665-1683







'he saw me at my uncle's'



'he saw me at my uncle's'

Middle Polish Dependency Treebank (MPDT)

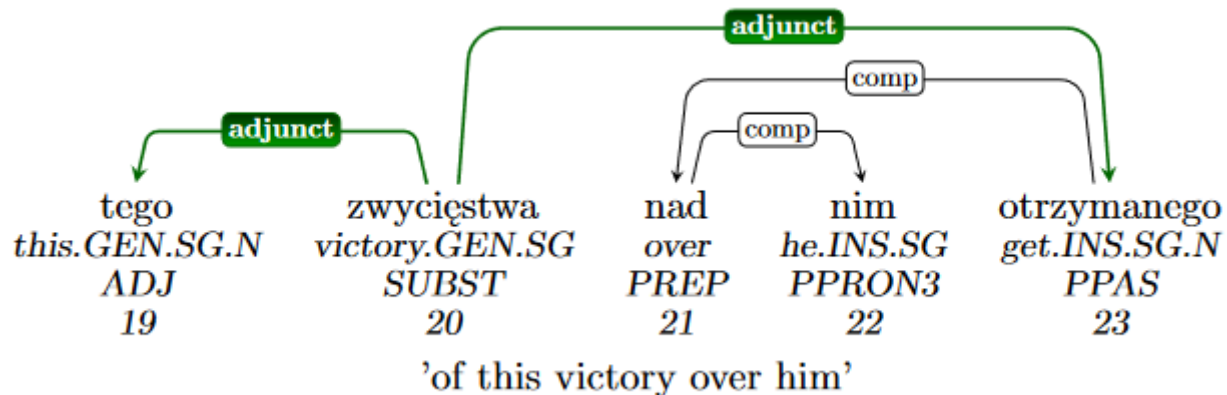
- 1,015 sentences (23,242 tokens)
 - dependency annotation based on Polish Dependency Bank, 22,200 sentences (<http://zil.ipipan.waw.pl/PDB>; Wróblewska & Rybak 2019)
 - sentences from 10 to 50 tokens, 22.9 tokens per sentence on average (15.8 in PDB)
 - format: CONLL and UD (in preparation)
-
- Author: Aleksandra Wieczorek
 - Consultations: Alina Wróblewska
 - Annotation and IT work: Bożena Itoya, Dorota Komosińska, Emanuel Modrzejewski
 - Funding: Polish Ministry of Science and Higher Education, National Programme for the Development of Humanities grant (project number 0413/NPRH7/H11/86/2018)

Electronic Corpus of the 17th- and 18th-century Polish Texts

- Cryptonym: KORBA (KORpus BARokowy ‘baroque corpus’)
- 13.5 M tokens (25 M tokens in autumn 2023)
- Principal investigator: Włodzimierz Gruszczyński
- richly annotated
- rich metadata

- Funding: Polish Ministry of Science and Higher Education, National Programme for the Development of Humanities grant (project number 0036/NPRH2/H11/81/2012 and 0413/NPRH7/H11/86/2018)
- Project duration: 2013-2018 and 2019-2023
- Coordinating body: Institute of Polish Language, Polish Academy of Sciences
- Cooperation: Institute of Computer Science, Polish Academy of Science

Methodology



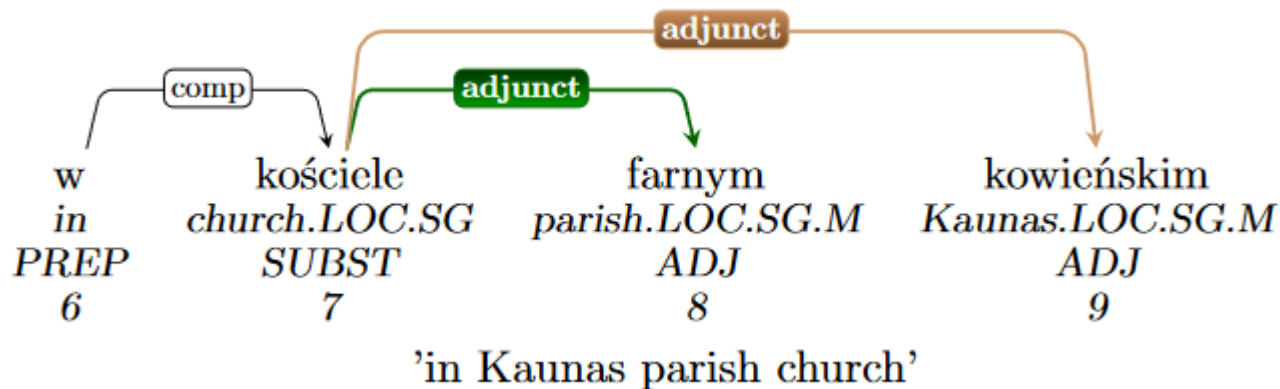
The distance between the noun and AM:

zwycięstwa → tego: -1

zwycięstwa → otrzymanego: 3

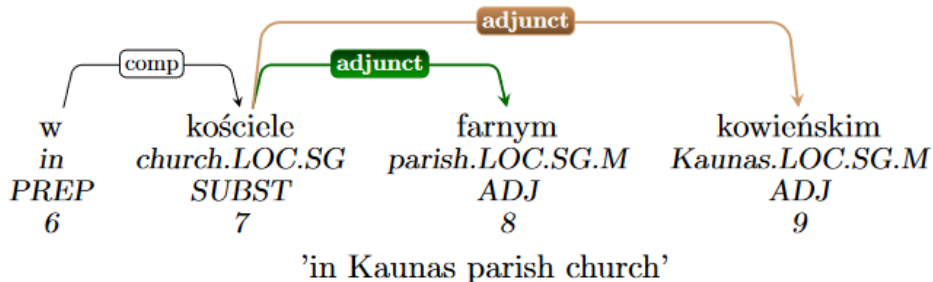
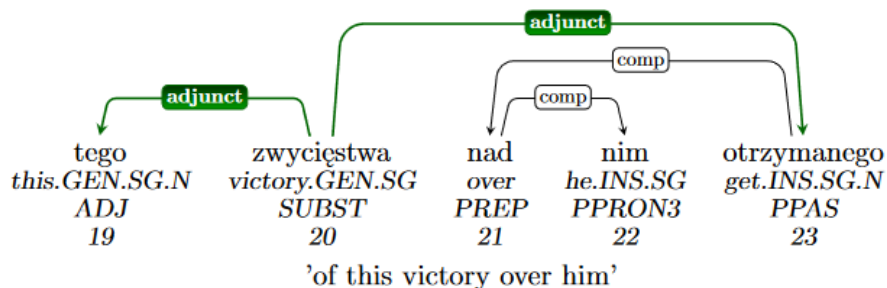
Methodology

Cases where there is another AM, an apposition, or a punctuation mark between the noun and AM are excluded from the calculations.



An excerpt from a table containing AMs from MPDT

Sent. ID	Sent. length	Governor	Modifier	Gov. POS	Mod. POS	Gov. number	Mod. number	Dist.	Abs. dist.
422	28	zwycięstwa	tego	SUBST	ADJ	20	19	-1	1
422	28	zwycięstwa	otrzymanego	SUBST	PPAS	20	23	3	3
546	10	kościół	farnym	SUBST	ADJ	7	8	1	1
...									



Results — general

- 2,156 AMs in 1,015 sentences
- 2.1 AM per sentence on average
- 150 sentences with no AM

Results — the distance between a noun and AM

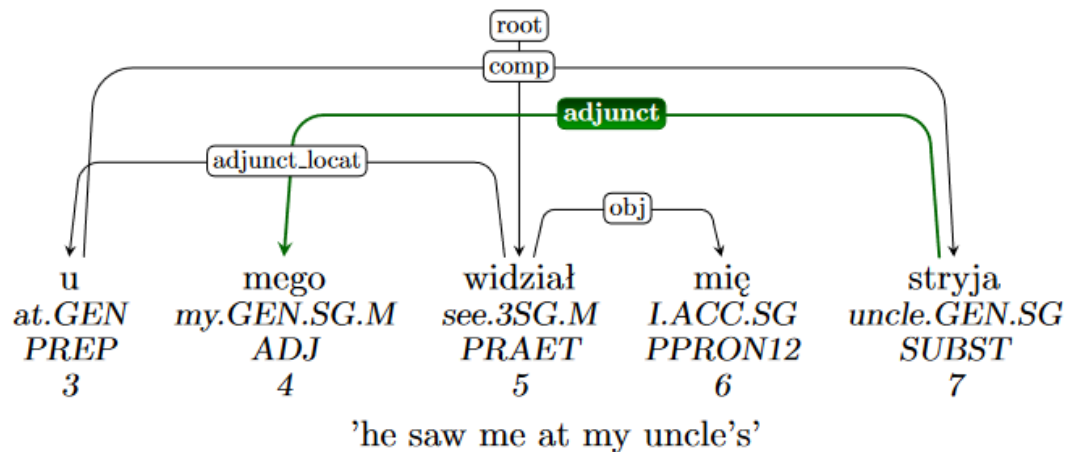
- the number of discontinuous noun phrases with AMs (distance more than 1): **391** (17% of all noun+AM pairs)
- the largest distance between noun and AM: **16**
- the average distance between noun and AM: **1.56**

Results – left and right position of AM

- the number of AMs on the left: 1,157
 - ***każdy** szlachcic*
every nobleman
- the number of AMs on the right: 1,123
 - *dygresje* **częste**
digressions **frequent**

Results – non-projective trees

	Middle Polish	Modern Polish
Crossing edges	8.9%	1.8%
Non-projective trees	46.1%	8.6%



Results – POS of AM

distance	adjectives	% of adjectives	participles	% of participles
=1	1694	95,98	71	4,02
>2	148	61,92	91	38,08
>5	33	46,48	38	53,52
>8	7	29,17	17	70,83
>12	2	20,00	8	80,00
total	2094	91,84	186	8,16

Noun phrases with participles

- *człęk na ziemi* **siedzący**
man on the ground **sitting**
- *szańce od nieprzyjaciela* **postawione**
entrenchments by the enemy **put up**
- *batalia piórem* **rysowana**
a battle with a pen **drawn**

Further research

- comparison with the contemporary Polish (PDB)
- refinement of the methodology
- research on the position of AMs taking into account the metadata

Other plans for the future

- expanding the Middle Polish Dependency Treebank to 2,000 sentences

Thank you!

aleksandra.wieczorek@ijp.pan.pl

Selected bibliography

- Gruszczyński, W., Adamiec, D., Bronikowska, R., Kieraś, W., Modrzejewski, E., Wieczorek, A., and Woliński, M. (2022). [*The Electronic Corpus of 17th- and 18th-century Polish Texts*](#). *Language Resources and Evaluation* 56:309-332.
- Karplukówna, M. (1967). *Język Jana Kurkowskiego (około 1580-1639)*. Wrocław - Warszawa - Kraków.
- Mel'čuk, I. A. (1988). *Dependency syntax: theory and practice*. Albany: SUNY Press.
- Perczyńska, N. (1978). *Uwagi o kolokwializmach składniowych w języku pamiętników Jana Chryzostoma Paska*. [In] J. Twardzikowa (ed.). *Studia z polskiej składni historycznej* v. II. Wrocław, pp. 59-64.
- Śliwiński, W. (2006). [*Inwersje barokowe w Kanikule i Lutni Jana Andrzeja Morsztyna*](#). *LingVaria* 1(1). Kraków: Wydział Polonistyki Uniwersytetu Jagiellońskiego, pp. 103-116.
- Topolińska, Z. (1984). *Składnia grupy imiennej*. [In] Z. Topolińska (ed.). *Gramatyka współczesnego języka polskiego. Składnia*, Warszawa, pp. 301-389.
- Wiśniewska, H. (1994). *Kulturalna polszczyzna XVII wieku. Na przykładzie Zamościa*. Lublin.
- Wróblewska, A., and Rybak, P. (2019). Dependency parsing of Polish. *Poznań Studies in Contemporary Linguistics*, 55(2):305-337.
- Wyderka, B. (2005). *Jak opisywać i interpretować składnię staropolskiej prozy retorycznej*. [In] S. Borawski (ed.). *Rozprawy o historii języka polskiego*. Zielona Góra, pp. 423-467.