

# Integration of the Electronic Dictionary of the 17th-18th c. Polish and the Electronic Corpus of the 17th and 18th c. Polish Texts

Renata Bronikowska<sup>1</sup>, Zbigniew Gawłowicz<sup>2</sup>,  
Maciej Ogrodniczuk<sup>2</sup>, Aleksandra Wieczorek<sup>1</sup> and Mateusz Żółtak<sup>3</sup>

1. Institute of Polish Language, Polish Academy of Sciences
2. Institute of Computer Science, Polish Academy of Sciences
3. Austrian Centre for Digital Humanities

# e-SXVII (the Electronic Dictionary of the 17th-18th c. Polish)

- Available online since 2004
- Planned size: ca 80K entries
- Now contains ca 39K entries (including ca 26K “germ” entries)
- Ca 77K forms – paradigms with blanks  
(only forms found in examined sources)
- Transliterated quotations from ca 1100 sources

<https://sxvii.pl/>



# KorBa (the Electronic Corpus of the 17th and 18th c. Polish Texts)

- Available online since 2018
- Now contains 13.5M tokens (sources from 1601-1772)
- Planned size: 25M tokens (including ca 4M from 1773-1800)
- Transliteration and transcription
- Morphosyntactically annotated by two taggers:  
Toygger and Concraft

<https://www.korba.edu.pl>

# Corpus → dictionary

Zapytanie  
[base="skrzydło"]

á é Ā Ę

KONSTRUKTOR ZAPYTAŃ ODRZUĆ OBCE SEGMENTY METADANE ▾

Warstwa wyświetlania: transliterowana Liczba wyników na stronę: 10

Wyszukaj

Lp	Lewy kontekst	Znalezienie	Rezyultat	Skrót	Data
1	tu był w namieście. On to vsłyszawszy/ podniósł	<a href="#">skrzydła [skrzydło:subst:locn]</a>	namiotu/ y rzekł do nich. Prakać wam przydzie	BudnyBPow	1614
2	bojącym się imienia moiego stońce sprawiedliwości/ y zdrowie ná	<a href="#">skrzydłách [skrzydło:subst:pl:locn]</a>	iego. Orietur Stella ex Iacob, et Homo ex	KorRoz	1645

CLICK!

## ELEKTRONICZNY SŁOWNIK JĘZYKA POLSKIEGO XVII I XVIII WIEKU

### SKRZYDŁO

Część mowy: rzecz.

Warianty fonetyczne: SKRZYDŁO, KRZYDŁO

Znaczenia:

1. »jedna z dwóch kończyn ptaka służących do latania«
2. »boczne oddziały wojska ustawionego w szyku bojowym«

Status hasła: w opracowaniu

[Odknośnik do słownika](#)

CLICK!

- more grammatical and semantic information
- link to the dictionary

# Dictionary → corpus (more quotations) 1/2

SKRZYDŁO

rzecz.

n

W TRAKCIE OPRACOWANIA



Warianty fonetyczne: SKRZYDŁO, \*KRZYDŁO

▼ Notowanie w słownikach

▼ Formy gramatyczne

▼ Znaczenia

▼ 1. »jedna z dwóch kończyn ptaka służących do latania«

- – Zawsze od bystrey wody Sokół stroni Gdy Czuie w [krzydłach](#) że piora wyroni. [PasPam 51.](#)
- – Są y insze [ryby] cudowne bardzo co iak uptakow [skrzydła](#) iak ubocianow nosy y głowy. [PasPam 65.](#)

▼ 2. »boczne oddziały wojska ustawionego w szyku bojowym«

▼ Więcej cytatów w Korpusie Barokowym

**CLICK!**

# Dictionary → corpus (more quotations) 2/2

Zapytanie  
((base="skrzydło"&pos="subst"))((base="krzydło"&pos="subst"))

á é Á Ę

KONSTRUKTOR ZAPYTAŃ ODRZUĆ OBCE SEGMENTY METADANE ▾

Warstwa wyświetlania  
transliterowana ▾











Liczba wyników na stronie  
10 ▾

Wyszukaj

Znaleziono 500 wyników.

Lp	Lewy kontekst	Rezultat	Prawy kontekst	Skrót	Data
1	tu był w namieście. On to vsłyszawszy/ podniósł	<a href="#">skrzydła [skrzydło:subst:pl:acc:n]</a>	namiotu/ y rzekł do nich: Płakć wam przydzie	BudnyBPow	1614
2	bojącym się imienia mojego stońce sprawiedliwości/ y zdrowie ná	<a href="#">skrzydłach [skrzydło:subst:pl:loc:n]</a>	iego. Orietur Stella ex Iacob, et Homo ex	KorRoz	1645
3	desolationis erit super alas: obrzydenie spustoszenia będzie ná	<a href="#">skrzydłach [skrzydło:subst:pl:loc:n]</a>	. To oboie Prorok wyraża/ y Woysko przyszłego Hetmána	KorRoz	1645
4	tedy rozumieć się ma Dániel Prorok/ to iest rościągnione	<a href="#">skrzydła [skrzydło:subst:pl:nom:n]</a>	Orłá ná Chorągwiách Rzymskich widziáne w Zydowstwie/ pewnym znákiem	KorRoz	1645
5	Chrześciáńskich. Ani iednák przez to biały Orzeł opuścił swe	<a href="#">skrzydła [skrzydło:subst:pl:acc:n]</a>	/ bo y tym dopierano:ále (według Ezechiela	DembWyw	1633
6	pokazało przez náwiedzeni nas Boże Wiárą S. złożył swe	<a href="#">skrzydła [skrzydło:subst:pl:acc:n]</a>	Polski Orzeł. Iednákże nic mu przez to zacności	DembWyw	1633
7	/ że biały Orzeł nie długo znowu przez wszystek świat	<a href="#">skrzydła [skrzydło:subst:pl:acc:n]</a>	swe rościągnie/ gdy ktoryszkolwiek Krol Polski ábo Aquiloński/	DembWyw	1633
8	sromotnie odpłoszył. Iuż tenże nie dawno przed tem	<a href="#">skrzydła [skrzydło:subst:sg:gen:n]</a>	białego Orłá przez Moskwę rozciągnął. Iuż go wieści gniewać	DembWyw	1633
9	się przeto zewnetrzne dziurka o wpuszczone powietrze na	<a href="#">skrzydła [skrzydło:subst:pl:acc:n]</a>	owego małego wiatraku z ímpetem wpadając w koło go	ChrósSEizyka	1764

## Formy gramatyczne

- Ip M. skrzydło 
- D. skrzydła 
- C. skrzydłu 
- B. skrzydło 
- N. skrzydłem 
- Ms. skrzydle 
- Im M. skrzydła 
- B. skrzydła 
- Ms. krzydłach  || skrzydłach 

**CLICK!**

# Dictionary → corpus (grammatical forms)

Zapytanie

[orth="krzydłach"&pos="subst"&number="pl"&case="loc"]

á

é

Á

É

KONSTRUKTOR ZAPYTAŃ

ODRZUĆ OBCE SEGMENTY

METADANE ▾

Warstwa wyświetlania  
uwpółcześniona ▾

Liczba wyników na stronę  
10 ▾

Wyszukaj

Znaleziono 1 wyników.

Lewy kontekst

Rezultat

Prawy kontekst

Skrót

Data

ochoty zawsze od bystrej wody sokół stroni gdy czuje w

krzydłach [krzydło:subst:pl:loc:n]

że pióra wyroni. Nietakiec nasze miało być rozstanie nie

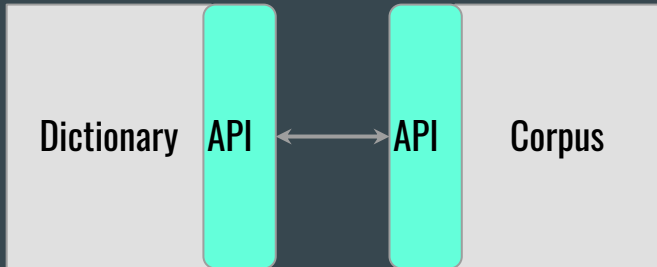
PasPam

między 1656 a 1688

# Dynamic vs static integration

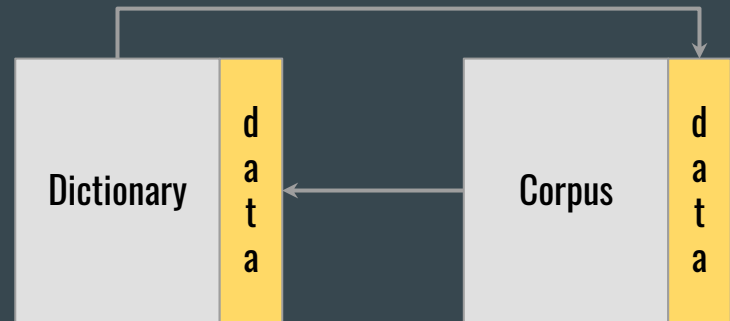
## Dynamic

- Add network APIs to both systems
- Query one system from the other **on demand**



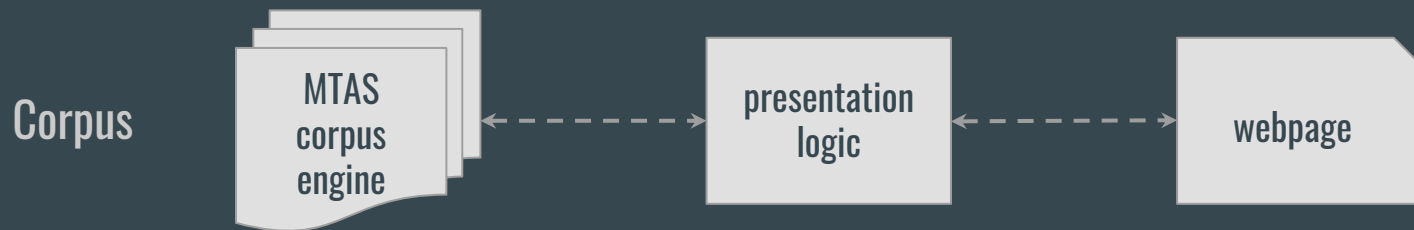
## Static

- Extend both systems databases
- **Batch-transfer** the data





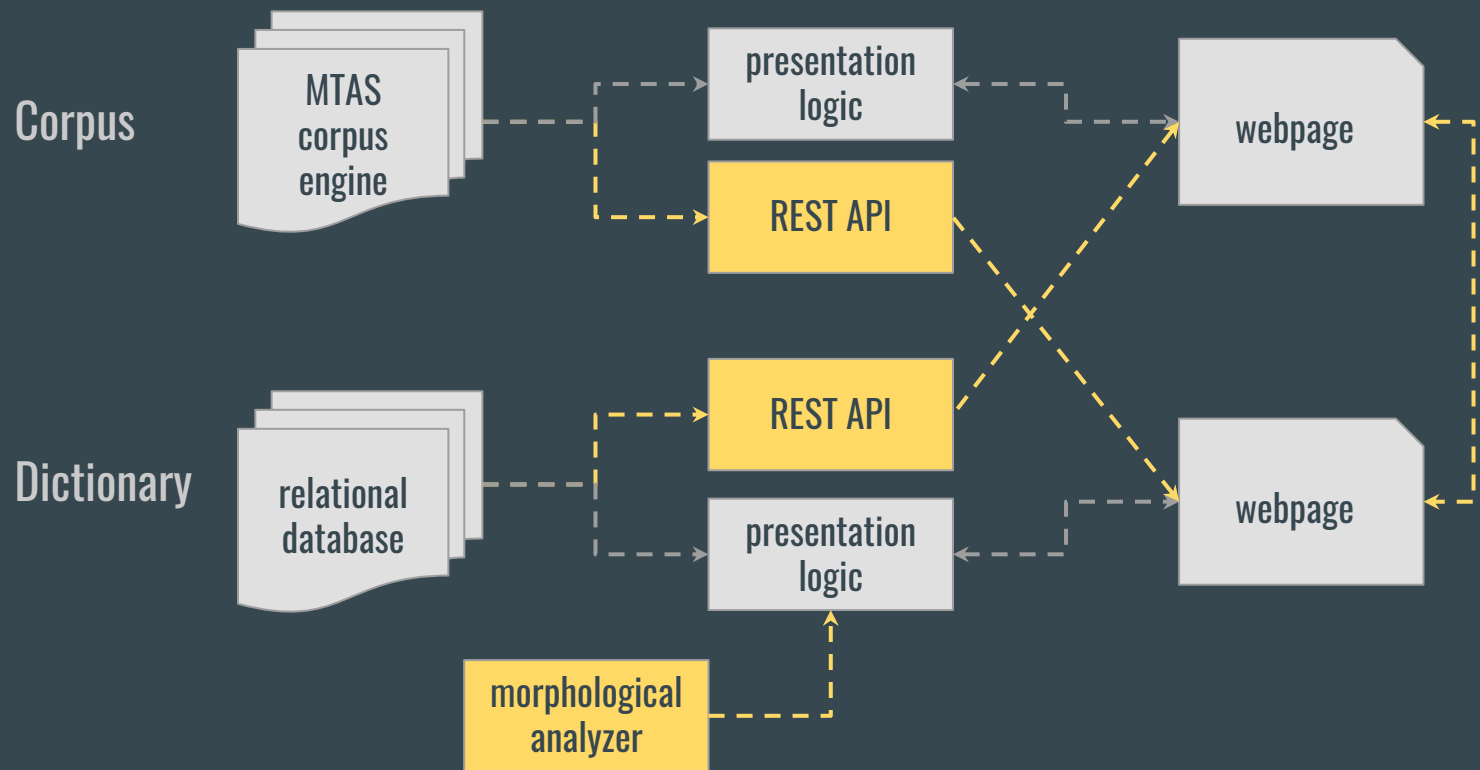
# Architecture - starting point



**Lack of network (REST) APIs!**



# Architecture - implemented solution



# Difficulties

## Complexity of the Polish grammar

- gerunds as nouns in the dictionary and as verbs in the corpus
- all the inflection forms of Polish verbs...
- etc.

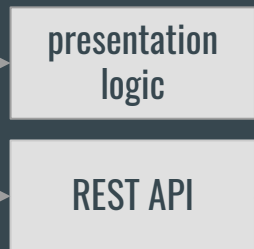
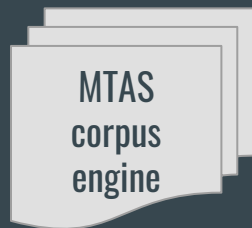
## Source text locations

- same source text described in a different way

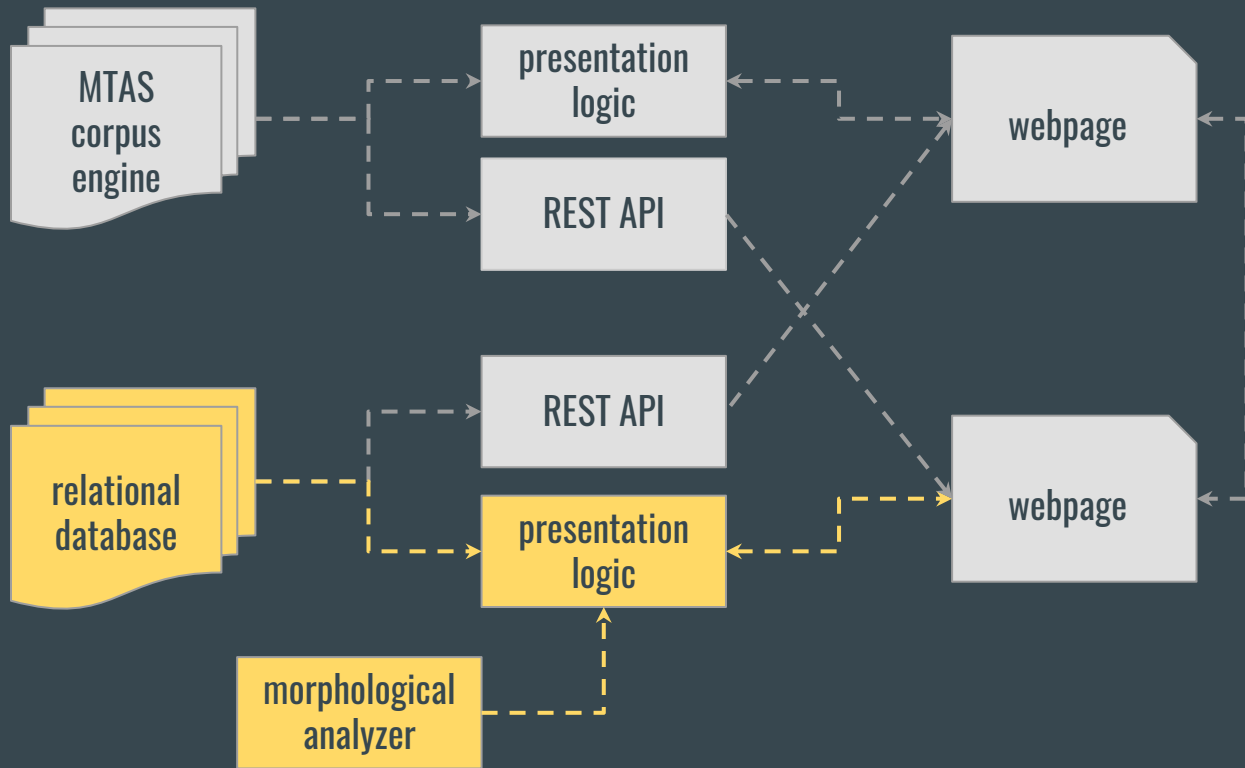
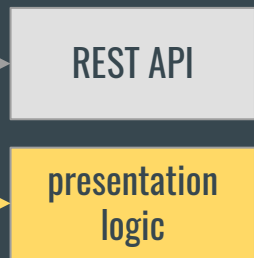
# Solutions - grammar

- some differences can be described by (relatively) **simple rules**
- **gerunds as nouns** in the dictionary and as **verbs in the corpus**
- we can **use morphological analyzer** used to tag the corpus on the dictionary side
- all the inflection forms of **Polish verbs...**

Corpus



Dictionary



# Solutions - grammar - gerund example

*PISANIE* 'writing' (*subst*)

1. Normalize - *pisanie*
2. Analyze
  - a. *pisanie*, subst, sg:acc:n
  - b. *pisanie*, subst, sg:nom:n
  - c. *pisanie*, subst, sg:nom:n
  - d. *pisanie*, subst, sg:voc:n,
  - e. *писа́ć*, *ger*, sg:nom.acc:n:imperf:aff
3. Matches the rule  
*if there is at least one pos=ger, use  
base forms noted by the dictionary with pos=subst  
and base forms resulting from the analyze where  
pos=ger with pos=ger*
4. The right corpus query for the whole dictionary entry is ([link](#))  
([base="pisanie"&pos="subst"])  
| ([base="писа́ć"&pos="ger"])

# Solutions - grammar - gerund example

*PISANIE* 'writing' (*subst*)

1. Normalize - *pisanie*
2. Analyze
  - a. *pisanie*, *subst*, *sg:acc:n*
  - b. *pisanie*, *subst*, *sg:nom:n*
  - c. *pisanie*, *subst*, *sg:nom:n*
  - d. *pisanie*, *subst*, *sg:voc:n*,
  - e. *писа́ć*, *ger*, *sg:nom.acc:n:imperf:aff*
3. Matches the rule  
*if there is at least one pos=ger, use  
base forms noted by the dictionary with pos=subst  
and base forms resulting from the analyze where  
pos=ger with pos=ger*
4. The right corpus query for the whole dictionary entry is ([link](#))  
[[base="pisanie"&pos="subst"]]  
| [[base="писа́ć"&pos="ger"]]

# Solutions - grammar - gerund example

*PISANIE* 'writing' (*subst*)

1. Normalize - *pisanie*

2. Analyze

a. *pisanie*, *subst*, *sg:acc:n*

b. *pisanie*, *subst*, *sg:nom:n*

c. *pisanie*, *subst*, *sg:nom:n*

d. *pisanie*, *subst*, *sg:voc:n*,

e. *pisać*, *ger*, *sg:nom.acc:n:imperf:aff*

3. Matches the rule

*if there is at least one pos=ger, use*

*base forms noted by the dictionary with pos=subst*

*and base forms resulting from the analyze where*

*pos=ger with pos=ger*

4. The right corpus query for the whole dictionary entry is ([link](#))

`([base="pisanie"&pos="subst"])`

`| ([base="pisać"&pos="ger"])`





# Solutions - grammar - gerund example

## *PISANIE* 'writing' (*subst*)

1. Normalize - *pisanie*
2. Analyze
  - a. *pisanie*, subst, sg:acc:n
  - b. *pisanie*, subst, sg:nom:n
  - c. *pisanie*, subst, sg:nom:n
  - d. *pisanie*, subst, sg:voc:n,
  - e. *pisać*, ger, sg:nom.acc:n:imperf:aff
3. Matches the rule  
*if there is at least one pos=ger, use base forms noted by the dictionary with pos=subst and base forms resulting from the analyze where pos=ger with pos=ger*
4. The right corpus query for the whole dictionary entry is ([link](#))  
[[base="pisanie"&pos="subst"]]  
| ([base="pisać"&pos="ger"])]
5. Similar rule for particular inflectional forms  
*if there is at least one pos=ger, override the pos condition with subst/ger*  
e.g. *pisania* (number: sg, case: gen) ([link](#))  
[orth="pisania"&number="sg"&case="gen"  
&pos="ger|subst"  
]

# Solutions - grammar - verb with “się” example 1/2

*ŚMIAĆ SIĘ* ‘to laugh’ (*reflexive verb*)

1. Normalize - *śmiać + się*

2. Analyze

a. śmiać, inf, imperf

3. Matches the rule

*If it's a verb, use base forms*

*resulting from the analyze*

*with a `pos!=ger/ppas/pact/ppraet` filter*

resulting in

`[base="śmiać"&!pos="ger/ppas/pact/ppraet"]`

# Solutions - grammar - verb with “się” example 1/2

## ŚMIAĆ SIĘ ‘to laugh’ (*reflexive verb*)

1. Normalize - *śmiać + się*

2. Analyze

a. śmiać, inf, imperf

3. Matches the rule

*If it's a verb, use base forms  
resulting from the analyze*

*with a `pos!=ger|ppas|pact|ppraet` filter*

resulting in

`[base="śmiać"&!pos="ger|ppas|pact|ppraet"]`

4. Apply the “się” rule

*“się” is represented by `[base="się"&pos="qub"]`  
before or after the main search term*

resulting in [\(link\)](#)

```
(  
  [base="się"&pos="qub"] ←  
  [base="śmiać"&!pos="ger|ppas|pact|ppraet"]  
) | (  
  [base="śmiać"&!pos="ger|ppas|pact|ppraet"]  
  [base="się"&pos="qub"] ←  
)
```

# Solutions - grammar - verb with “się” example 2/2

*ŚMIAĆ SIĘ* (reflexive verb) - *śmiali się* ‘they laughed’ form

1. Normalize - *śmiali* + *się*

2. Analyze

- a. śmiać, fin, sg:ter:imperf + li, qub
- b. śmiały, adj, pl:nom.voc:manim1.p1
- c. śmiać, praet, pl:manim1.p1:imperf

3. Doesn't match any rule -  
use whole analyze results

```
[base="śmiały"&orth="śmiali"&degree="pos"&number="pl"&pos="adj"]  
| [base="śmiać"&orth="śmiali"&aspect="imperf"&number="pl"&pos="praet"]  
| [base="śmiać"&orth="śmia"&aspect="imperf"&number="sg"  
&person="ter"&pos="fin"] [base="li"&orth="li"&pos="qub"]
```

# Solutions - grammar - verb with “się” example 2/2

*ŚMIAĆ SIĘ* (reflexive verb) - *śmiali się* ‘they laughed’ form

1. Normalize - *śmiali* + *się*

2. Analyze

- a. śmiać, fin, sg:ter:imperf + li, qub
- b. śmiały, adj, pl:nom.voc:manim1.p1
- c. śmiać, praet, pl:manim1.p1:imperf

3. Doesn't match any rule -  
use whole analyze results

[base="śmiały"&orth="śmiali"&degree="pos"&number="pl"&pos="adj"] ←

| [base="śmiać"&orth="śmiali"&aspect="imperf"&number="pl"&pos="praet"] ←

| [base="śmiać"&orth="śmia"&aspect="imperf"&number="sg"  
&person="ter"&pos="fin"] [base="li"&orth="li"&pos="qub"] ←

# Solutions - grammar - verb with “się” example 2/2

*ŚMIAĆ SIĘ (reflexive verb) - śmiali się* ‘they laughed’ form

1. Normalize - *śmiali + się*

2. Analyze

- a. śmiać, fin, sg:ter:imperf + li, qub
- b. śmiały, adj, pl:nom.voc:manim1.p1
- c. śmiać, praet, pl:manim1.p1:imperf

3. Doesn't match any rule -  
use whole analyze results

```
[base="śmiały"&orth="śmiali"&degree="pos"&number="pl"&pos="adj"]  
| [base="śmiać"&orth="śmiali"&aspect="imperf"&number="pl"&pos="praet"]  
| [base="śmiać"&orth="śmia"&aspect="imperf"&number="sg"  
&person="ter"&pos="fin"]|[base="li"&orth="li"&pos="qub"]
```

4. Apply the “się” rule  
“*się*” is represented by  
*[base="się"&pos="qub"]* before or after  
the main search term

resulting in ([link](#))  
...too long to be shown...

# Solutions - grammar - verb with “się” example 2/2

*ŚMIAĆ SIĘ (reflexive verb) - śmiali się* ‘they laughed’ form

1. Normalize - *śmiali + się*

2. Analyze

- a. śmiać, fin, sg:ter:imperf + li, qub
- b. śmiały, adj, pl:nom.voc:manim1.p1
- c. śmiać, praet, pl:manim1.p1:imperf

3. Doesn't match any rule -  
use whole analyze results

```
[base="śmiały"&orth="śmiali"&degree="pos"&number="pl"&pos="adj"]  
| [base="śmiać"&orth="śmiali"&aspect="imperf"&number="pl"&pos="praet"]  
| [base="śmiać"&orth="śmia"&aspect="imperf"&number="sg"  
&person="ter"&pos="fin"]|[base="li"&orth="li"&pos="qub"]
```

4. Apply the “się” rule  
“*się*” is represented by  
*[base="się"&pos="qub"]* before or after  
the main search term

resulting in ([link](#))  
...too long to be shown...

but fetches  
the right results

# Plans for the future

Dictionary and corpus integration for the dictionary editors (in progress):

- automatic download of the inflectional forms
- automatic download of the quotations

Integration of other resources:

- The Digital Library of Polish and Poland-related Ephemeral Prints from the 16th, 17th and 18th Centuries
- card index of the Electronic Dictionary of the 17th-18th Century Polish





# Thank you!

[renata.bronikowska at ijp.pan.pl](mailto:renata.bronikowska@ijp.pan.pl)

[zbigniew.gawlowicz at ipipan.waw.pl](mailto:zbigniew.gawlowicz@ipipan.waw.pl)

[maciej.ogrodniczuk at ipipan.waw.pl](mailto:maciej.ogrodniczuk@ipipan.waw.pl)

[aleksandra.wieczorek at ijp.pan.pl](mailto:aleksandra.wieczorek@ijp.pan.pl)

[mateusz.zoltak at boku.ac.at](mailto:mateusz.zoltak@boku.ac.at)