

Natalia Kotsyba  
Institute of Interdisciplinary Studies  
Warsaw University

**PolUKR**  
a Polish-Ukrainian Parallel Corpus  
as a testbed for parallel corpora  
toolbox

<http://corpus.domeczek.pl>

„SlaviCorp”, Warsaw University

23 November 2010

# Overview of the talk

- Objectives and history
- Content of the corpus
- Tools: UGTag, PLUczeK
- Searching through the corpus
- Current work and further developments

# Objectives

- rather quality than size oriented
- originally planned as a pilot project that had to serve for raising awareness of corpora as linguistic tools, especially for such an under-resourced language as Ukrainian
- preparing a methodological and instrumental background for do-it-yourself parallel corpora (currently some of the developed tools are used in at least three individual PhD projects)
- back in 2004 there were very few tools for creating parallel corpora, nowadays the situation is much better but still not ideal

# Time line

- 2004 - first texts acquired (2005 first public release, ca. 600 K tokens, paragraph alignment)
- November 2007- October 2009 the project was supported a grant by the Ministry of Science and Higher Education of Poland (2009 second release, morphologically annotated, sentence alignment, not public)
- after 2009 on (hopefully 2010 second public release)

# Contents and genre distribution

- size: ca. 3 mln. tokens
- not balanced, with a strong prevalence of fiction and more Ukrainian originals than Polish ones
- magazine articles: 9 %
- fiction: 85 %
- educational literature, textbooks: 5 %
- official correspondence: 0.3 %
- technical documentation: 0.7 %

# Benefits of the grant period

- purchasing modern literature (practically not available in libraries, also necessary for copyright issues)
- scanning, digitalizing and edition of texts
- developing and bringing to life a common morphosyntactic conceptual scheme
- developing user friendly tools for developing parallel corpora

# A common morphosyntactic conceptual scheme: Ukrainian

- data: Ukrainian Grammatical Dictionary with ca. 165K lemmas – acknowledgements to Igor V. Shevchenko and Volodymyr Shyrokov
- 384 tags used in the Ukrainian Grammatical Dictionary were extended to a MULTEXT-East format with ca. 1200 tags -- many categories were presented in a more detailed way, especially pronouns
- see example of an undisambiguated tagged text in the next slide

```
<?xml version="1.0"?>
<text id="aisecuk." lang="uk">
<body>
<p id="aisecuk.1">
<s id="aisecuk.1.1">
<w lemma="AIESEC" ana="X">AIESEC</w>
<c>(</c>
<w lemma="міжнародний" ana="Ao-fsns">міжнародна</w>
<w lemma="асоціація" ana="Ncfsnn">асоціація</w>
<w_>
<w lemma="студент" ana="Ncmpry">студентів</w>
<w lemma="студент" ana="Ncmprgy">студентів</w>
</w_>
<c>,</c>
<w_>
<w lemma="який" ana="Pq----pna">які</w>
<w lemma="який" ana="Pq---праа">які</w>
<w lemma="який" ana="Pr----pna">які</w>
<w lemma="який" ana="Pr---праа">які</w>
</w_>
<w_>
<w lemma="займатися" ana="Vmpip3p">займаються</w>
<w lemma="займатися" ana="Vmpip3p">займаються</w>
</w_>
<w lemma="економія" ana="Ncfsin">економією</w>
<w_>
<w lemma="й" ana="Ccs">й</w>
<w lemma="й" ana="Q">й</w>
</w_>
<w_>
<w lemma="управління" ana="Ncnpdn">управлінням</w>
<w lemma="управління" ana="Ncnsin">управлінням</w>
</w_>
<c>)</c>
```

# A common morphosyntactic conceptual scheme: Polish

- ca. 1000 IPIAN tag configurations reconceptualized into ca. 1200 into the MULTEXT-East mode.
- only half of the original tags found one-to-one projections, the rest are either results of contractions or more detailed representations
- trade-off: avoiding information loss vs fitting the previously designed MULTEXT-East conceptualization.
- see example of a disambiguated tagged text in the next slide

<text id="aisecpl" lang="pl">  
<body>  
<p id="aisecpl.1">  
<s id="aisecpl.1.1">  
<w lemma="aiesec" ana="N-mnnsn">AIESEC</w>  
<c></c>  
<w lemma="międzynarodowy" ana="A-pn--sn">międzynarodowe</w>  
<w lemma="stowarzyszenie" ana="Ngnn-snen">stowarzyszenie</w>  
<w lemma="student" ana="N-myypg">studentów</w>  
<w lemma="ekonomia" ana="N-f--sg">ekonomii</w>  
<w lemma="i" ana="C">i</w>  
<w lemma="zarządzanie" ana="N-n--sg">zarządzania</w>  
<c></c>  
<w lemma="to" ana="Pd--nnnsn--n">to</w>  
<w lemma="niepolityczny" ana="A-pf--sn">niepolityczna</w>  
<c>,</c>  
<w lemma="naukowy" ana="A-pf--sn">naukowa</w>  
<w lemma="i" ana="C">i</w>  
<w lemma="niekomercyjny" ana="A-pf--sn">niekomercyjna</w>  
<w lemma="organizacja" ana="N-f--sn">organizacja</w>  
<c>,</c>  
<w lemma="w" ana="Spl">w</w>  
<w lemma="całość" ana="N-f--sl">całości</w>  
<w lemma="zarządzać" ana="Ap-f--sn-ppn">zarządzana</w>  
<w lemma="przez" ana="Spa">przez</w>  
<w lemma="student" ana="N-myypa">studentów</w>  
<c>.</c>  
</s>  
</p>

# A common morphosyntactic conceptual scheme

- Both the Ukrainian and Polish tagsets, together with sample lexicons\* are part of MULTEXT-East v.4 (released in May 2010).

\*Acquiring the lexicon for Polish is due to the courtesy of Marcin Woliński and Zygmunt Saloni.

The Polish MULTEXT-East package also includes a one book corpus of George Orwell's "1984".

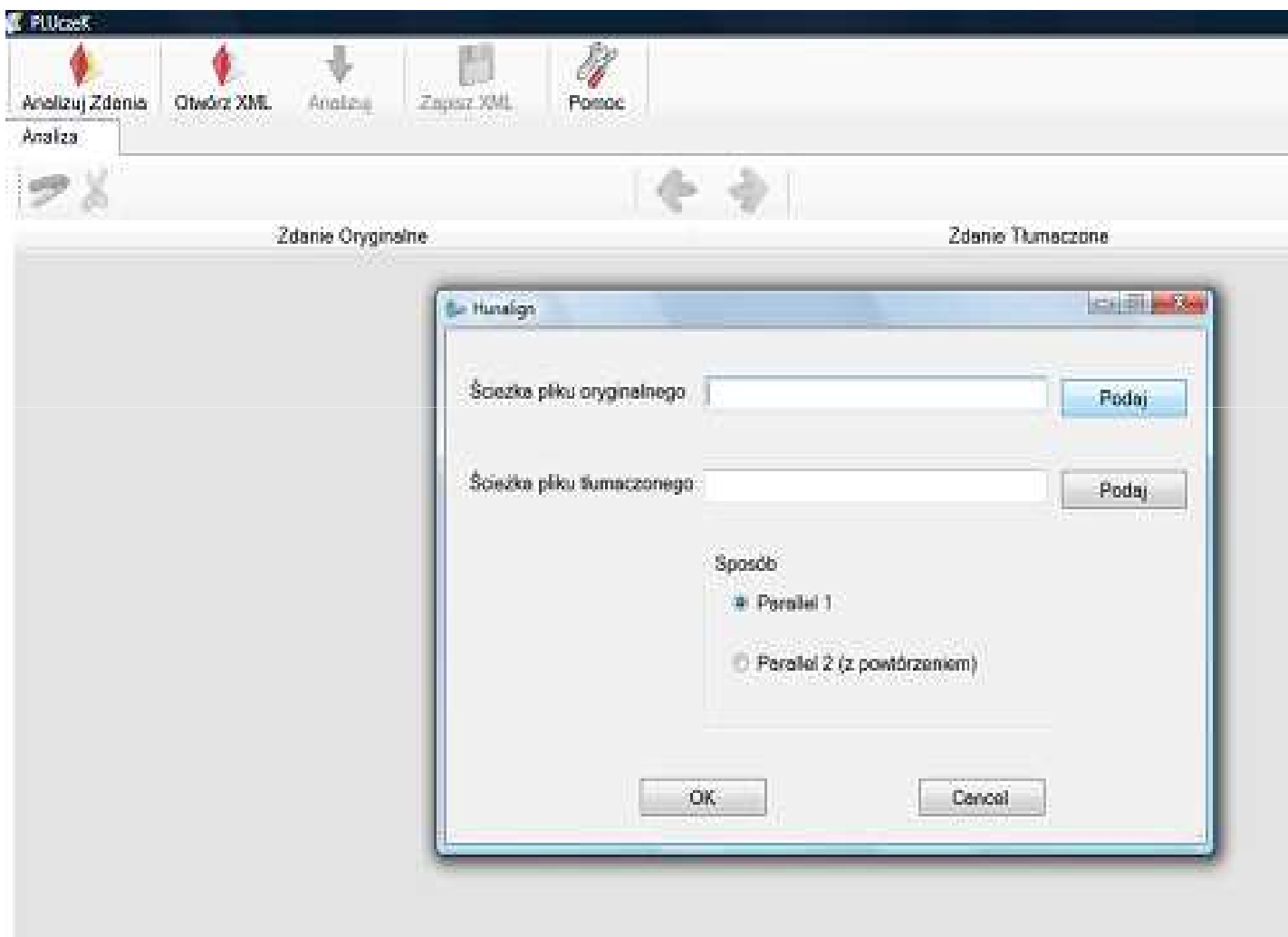
# Converters and other software

- TAKIPI → MULTEXT-East converter (Python + conversion tables for tags and lists selected of lemmas)
- PLUczeK, graphic editor and converter for alignments (C#)
- corpus manager with integrated modules: sentence splitter, a Ukrainian grammatical dictionary UGTag with support for entry edition, corpus maker (indexing and converting into a binary format), search engine (Java)

# Alignment with PLUczek

- PLUczek is a user friendly alignment editor and converter to XML
- written in C#, based in .NET platform
- uses Hunalign for pre-alignment
- tested in Windows XP Professional, Vista and Windows 7
- input: texts split into sentences and paragraphs encoded in UTF-8, .txt
- output: texts and stand-off alignment in XMLFormat

# PLUczek: pre-alignment with HunAlign



# PLUczek: working with prealigned texts

Zdanie Originalne	Zdanie Transkrypcyjne	Numerozacja
Stanisław Lem	Станіслав Лем	1 - 1
Solaris 5172	Соларіс 5103	2 - 2
Przybysz	ПРИБУТТЯ	3 - 3
O dziewiętnastej części pokładowego zszedłem, mijając stojących wokół studni, po metalowych szczeblach do wnętrza zasobnika.	Рівно о дев'ятнадцятій за бортовим часом я пройшов повз тих, що зібраніся довкола колодязя, й металевим трюпом опустився в капсулу.	4 - 4
Było w nim akurat tyle miejsca, aby umieścić łokcie.	Бона була така твора, що в ній ледве можна було розстелити піджак! Отак коли я вернувся наконечник шланга в патрубку, який спирався на стіну залуци і наї скафандра роздубав, я вже не міг навіть пророзумітися.	5 - 5,8
Po włączeniu końcówki w przewód, wystający ze ściany, skafander wydał się i odciążenie nie mogłem już wykonać najmniejszego ruchu.	Я став, – точніше висів, – у повітряному лажі, змещившись в одну цілу з металевим шкаралупом.	6 - 7
Stałem - czy raczej wisiałem - w powietrznym łożu, zespolony w jedną całość z metalową skorupą.	Зайшов очі, я побачив край опуклого скла стінки колодязя, а ялице – силікане над ним Моддардове обліччя.	7 - 8
Podniebieszcy oczy, zobaczyłem przez wypukłą szybę ściany studni i wyżej schyloną nad nią twarz Moddarda.	За мить воно щезло, і в капсулі зробилося темно – це негорі опустивши вогонь записний конус.	8 - 9
Znikła zaraz i zapadła ciemność, bo z góry nałożono ciężki ochronny szczek.	Вісім разів званли електромотори, які затягували гайки.	9 - 10
Słyszałem odmiaknięte powtórzone żwiast materów elektrycznych, która dosięgały śruby.	Пали змечало повітря, яке нагнітали в емуртаватори.	10 - 11
Potem - syk wpuszczonego do amortyzatorów powietrza.	Очі поступово занесли до темряви.	11 - 12
Wzrok przywykał do ciemności. Widziałem już zaledynowy kontur jedyne go wakażnika.	Я розрізняв уже темно-зелені обрисы таблиці – єдиного в кабіні.	12, 13 - 13
- Gotów, Kelvin? - rozległo się w słuchawkach.	- Ты готовий, Кельвіна? - пралучало в навушниках.	14 - 14
- Gotów, Moddard - odpowiedziałem.	- Готовий, Моддардо, - відповів я.	15 - 15
- Nie troszcz się o nic.	- Ні про що не турбуйся.	16 - 16
Stacja cię odbierze - powiedział.	Станція тебе прийме, - сказав він.	17 - 17
- Szczęśliwej drogi!	- Щесливої дороги!	18 - 18
Nim zdążyłem odpowiedzieć, coś zgrzytnęło w górze i zasobnik drgnął.	На встиг я відповісти, як негорі щось заскреготало, і капсула здригнулася.	19 - 19
Napięłam odruchowo mięśnie, lecz nic więcej się nie stało.	Я минають напружені м'язи, але вже більше нічого не відчула.	20 - 20
- Kody start? - spytałem i usłyszałem szmer, jakby ziarenka najdrobniejszego	- Коли старт? - запитав я і вловив якийсь шорох, немовби на мембрану.	21, 22 - 21

# PLUczek: editing alignments

The screenshot displays the PLUczek application window. At the top, there is a menu bar with icons for 'Analizuj Zdania', 'Otwórz XML', 'Anuluj', 'Zapisz XML', and 'Pomoc'. Below the menu bar, there are navigation arrows and a 'Zdania Oryginalne' / 'Zdania Tłumaczone' header. The main area contains a table with three columns: 'Zdania Oryginalne', 'Zdania Tłumaczone', and 'Numeracja'. A dialog box titled 'Zdanie do podziału' is open, showing a text input field with a highlighted segment of text and a 'Podziel' button. The table contains the following data:

Zdania Oryginalne	Zdania Tłumaczone	Numeracja
Nie miałem nic do roboty.	- Ти гадаєш.Я справді хочу тут залишитися.	4994 - 5205,5206
	СТАРИЙ ММОЦДЯ сидів біля великого вікна й споглядав Океан.	4995 - 5207,5208
	Роботи в мене не було жодної Ралари, складили за п'ять днів, тепер обернувся у тунель хитлі, що мчав у порожнечі десь поза суцільнім Океані.	4996 - 5209,5210
	Досить темної пиллової туманності, яка розпростерлася на обширі у всій трильйона кубічних миль і поглинає промені світла та будь-які інші сигнали, він потрапляє на перший із нескількох ретрансляторів.	4997 - 5211
	Звідти, від одного радіобуя до іншого, стрибками завдовжки в мільярд кілометрів він лінується по кривій величезної дуги, аж поки останній ретранслятор, металевий брїла, щільно навішований товченими приладами, з відраженням пісок направленої антени, сконцентрує його ще раз і скерує далі в простір, до Зеніт Потім мінуть місяці, і тріклячєко такий самий пучок енергії, посланий із Землі, протягши за собою борозну імпульсних спотворень у гравітаційному полі Галактики, досягне криві космічної зоряної прослизни, посланий; уздраж ланцюга був, що вільно дрейфують, і, не зменшуючи швидкості, пониніть до подвійного сонця планети Соларіс.Океан під високим чорним сонцем був чорніший, ніж звичайно.	4998 - 5212,5213,5214
	Рудий тунелі сталював його стики з обрисом.	4999 - 5215
	День був надзвичайно паркий, нби провіщав шалену бурю, яко кілька разів на рік потус на планеті.	5000 - 5216
	Були всі підстави припускати, що її єдиний мешканець контролює клімат і сам спричиняє ці бурі.	5001 - 5217
	Ця пару місяців мені доведеться взирати з цих вікон, з висоти спостерігати за розкутом білим залатом і стоніням багрянцем, як час від часу відбувалася в їхньому рідномууниверсенні, в сріблостому пухирі	5002 - 5218

# Stand-off alignment format

```
<?xml version="1.0" encoding="utf-8"?>
<cesAlign version="4.1">
  <linkList id="lemsolpluk">
    <linkGrp id="lemsolpl.1" type="body" targType="s" domains="lemsolpl lemsoluk">
      <link xtargets="lemsolpl.1.1 ; lemsoluk.1.1"/>
      <link xtargets="lemsolpl.1.2 ; lemsoluk.1.2"/>
      <link xtargets="lemsolpl.1.3 ; lemsoluk.1.3"/>
      ...
      <link xtargets="lemsolpl.8.15 ; lemsoluk.9.14"/>
      <link xtargets="lemsolpl.9.1 lemsolpl.9.2 ; lemsoluk.10.1"/>
      <link xtargets="lemsolpl.9.3 ; lemsoluk.10.2"/>
      <link xtargets="lemsolpl.10.1 ; lemsoluk.11.1"/>
      <link xtargets="lemsolpl.10.2 ; lemsoluk.11.2 lemsoluk.11.3"/>
      <link xtargets="lemsolpl.10.3 lemsolpl.10.4 ; lemsoluk.11.4 lemsoluk.11.5"/>
      ...
    </linkGrp>
  </linkList>
</cesAlign>
```

# Search options

- POSHUK -- the bundle of the corpus maker and corpus search modules (only very basic functions are realized, lots of bugs), one of possible development scenarios is moving it to sourceforge
- current local search support – ParaConc (conversion enabled by Intertext developed within InterCorp project)
- in the process of implementing PARK+Manatee
- ? CWB + PARASOL interface

AND

[tag="Spsa"]

## ALIGNMENT

tylko o stagnację . Rozmiary  
 nastąpił w 1992 r. , gdy hiperinflacja i gwałtowny spadek produkcji doprowadziły do załamania syst...  
 nastąpił w 1992 r. , gdy hiperinflacja i gwałtowny spadek produkcji doprowadziły do załamania syst...  
 rzypomnę , że Wielki Kryzys początku lat trzydziestych spowodował spadek produkcji na Zachodzie ...  
 ;podarce produkcja spadła natomiast ponad trzykrotnie . Rozmiary  
 ań sił produkcyjnych porównywalne są ze skutkami wojny i okupacji niemieckiej : gospodarka narod...  
 ań sił produkcyjnych porównywalne są ze skutkami wojny i okupacji niemieckiej : gospodarka narod...  
 tczeń potrzeba było pięciu lat ; w 1950 r . poziom przedwojenny został nawet przekroczony o 15 % ....  
 a zaznała jeszcze większych strat : pierwsza wojna światowa , rewolucja i wojna domowa 1917-192...  
 5-1926 r . , po pięciu-sześciu latach , przedwojenny poziom produkcji zasadniczo został przywrócony...  
 5-1926 r . , po pięciu-sześciu latach , przedwojenny poziom produkcji zasadniczo został przywrócony...  
 5-1926 r . , po pięciu-sześciu latach , przedwojenny poziom produkcji zasadniczo został przywrócony...  
 5-1926 r . , po pięciu-sześciu latach , przedwojenny poziom produkcji zasadniczo został przywrócony...  
 a Ukraina dopiero w zeszłym roku , po dziesięciu latach od początku jawnej stagnacji , osiągnęła ni...  
 a Ukraina dopiero w zeszłym roku , po dziesięciu latach od początku jawnej stagnacji , osiągnęła ni...  
 ów ić o średnich zarobkach i emeryturach , poziomie bezrobocia , spożyciu białka , gruźlicy , żebrak...  
 ów ić o średnich zarobkach i emeryturach , poziomie bezrobocia , spożyciu białka , gruźlicy , żebrak...  
 zpieczeństwo kryzysów polega na tym , że powodują one nieodwracalne uproszczenie systemu . Pr...  
 l za naukowcami z Kijowskiego Międzynarodowego Instytutu Socjologii podział społeczeństw wedłu...  
 l za naukowcami z Kijowskiego Międzynarodowego Instytutu Socjologii podział społeczeństw wedłu...  
 l za naukowcami z Kijowskiego Międzynarodowego Instytutu Socjologii podział społeczeństw wedłu...  
 l za naukowcami z Kijowskiego Międzynarodowego Instytutu Socjologii podział społeczeństw wedłu...  
 uropejskich na przestrzeni ostatniego dziesięciolecia widoczne było przejście od stadium industrial...  
 uropejskich na przestrzeni ostatniego dziesięciolecia widoczne było przejście od stadium industrial...  
 uropejskich na przestrzeni ostatniego dziesięciolecia widoczne było przejście od stadium industrial...  
 uropejskich na przestrzeni ostatniego dziesięciolecia widoczne było przejście od stadium industrial...  
 lat temu znajdowała się w stadium industrialno - agrarnym , a przez dziesięć lat przebudowy i refor...  
 lat temu znajdowała się w stadium industrialno - agrarnym , a przez dziesięć lat przebudowy i refor...

## LEFT CONTEXT

## MATCH

ж економіка скоротила виробництво біль... в [Spsa] [Spsg] [Spsl]  
 наслідками війни і нацистської окупації : в [Spsa] [Spsg] [Spsl]  
 народному господарстві визволеної від оку... в [Spsa] [Spsg] [Spsl]  
 . довоєнний рівень було навіть перевищено на [I] [Q] [Spsa] [Spsl]  
 рр . скоротили продуктивні сили України на [I] [Q] [Spsa] [Spsl]  
 дев'ять десятих . Проте , уже в [Spsa] [Spsg] [Spsl]  
 в 1925 - 1926 рр . , через [Spsa]  
 років , довоєнний рівень промисловості був в [Spsa] [Spsg] [Spsl]  
 сільське господарство ще раніше було пос... на [I] [Q] [Spsa] [Spsl]  
 Постсоціалістична Україна тільки минулог... через [Spsa]  
 дала невеликий ріст ( 6 % за [Q] [Rp] [Spsa] [Spsl] [Vmen]  
 згадувати . Не варто також говорити про [Spsa]  
 поширення туберкульозу , жебраків , прост... на [I] [Q] [Spsa] [Spsl]  
 , огрублення . Найбільша небезпека криз в [Spsa] [Spsg] [Spsl]  
 рівня складності системи . Приймемо спід... за [Q] [Rp] [Spsa] [Spsl] [Vmen]  
 Київського міжнародного інституту соціоло... за [Q] [Rp] [Spsa] [Spsl] [Vmen]  
 суспільств за макрогалузевою структурою (... за [Q] [Rp] [Spsa] [Spsl] [Vmen]  
 тобто за структурою зайнятості ) населення на [I] [Q] [Spsa] [Spsl]  
 стадії еволюції . В європейських націях за [Q] [Rp] [Spsa] [Spsl] [Vmen]  
 спостерігався рух від індустріально-інформ... на [I] [Q] [Spsa] [Spsl]  
 якій найбільша частина суспільної праці в... на [I] [Q] [Spsa] [Spsl]  
 - до інформаційно-індустріальної стадії , де на [I] [Q] [Spsa] [Spsl]  
 . Україна десять років тому знаходилась на [I] [Q] [Spsa] [Spsl]  
 знаходилась на стадії індустріально-аграр... за [Q] [Rp] [Spsa] [Spsl] [Vmen]  
 до аграрно-індустріальної стадії . Так , в [Spsa] [Spsg] [Spsl]  
 Так , в 1990 г . в [Spsa] [Spsg] [Spsl]  
 30 , 7 % населення ) , в [Spsa] [Spsg] [Spsl]  
 , 7 % ) . В 1999 в [Spsa] [Spsg] [Spsl]  
 ( 20 , 2 % ) в [Spsa] [Spsg] [Spsl]

## Meta-Information

## Wide Context

## Meta-Information

Name

Open Corpus ...

Close Corpus

Save Results ...

Search

Corpus Information

Preferences ...

Exit

Help

About

**Corpus Information**

General | **Text Properties**

Parallel Texts: Інформація про інформаційно-дослідн...

Monolingual Texts: Інформація про інформаційно-дослідн...

Title: дослідницький центр «Інтеграція та розвиток»

Description:

Author List:

Style: newsletter

Medium: file\_doc

Publisher: nd

Place:

Creation Date:

Publish Date: 1996,2004

First Publish Date:

Language: uk

Author Language: uk

Quality: 0

Availability: 1

When Added: Jan 27, 2010

Close Help

Ukrainian sub-query

AND [tag="Rp"]

	LEFT_CONTEXT	MATCH	
stosunku ...	до різних виборчих систем , а	також [Rp]	опитуван
czne na Uk...	різні науково-практичні заходи на У...	за [Q] [Rp] [Spsa] [Spsi] [Vmen]	її межах
nych . Kier...	свої завдання , передбачені статуто...	за [Q] [Rp] [Spsa] [Spsi] [Vmen]	рахунок г
	, єдності і співробітництву . Головні	напрямки [Rp] [Ncmpran] [Ncmprnn] [...]	діяльнос
	і діяльність - такий девіз завжди	разом [Ncmsin] [Rp]	з нами . (
nu wkrótce...	детальний опис збираємося послят...	часом [Ncmsin] [Rp]	. Перед с
acji na Ukr...	Перед семінаром відбудеться зустр...	після [Rp] [Spsg]	виборів "
acji na Ukr...	, присв'ячена аналізу ситуації в Укр...	після [Rp] [Spsg]	парламе
le wyników ...	Вас виступити в обговоренні „ Украї...	після [Rp] [Spsg]	виборів "
na Unia Eur...	квітня ). Мовами зустрічі „ Україна	після [Rp] [Spsg]	виборів "
etnia . Będ...	зустрічі та список учасників . Повідо...	також [Rp]	, що орга
pospolitej ...	так само , як це сталося	насправді [Rp]	. Поділяк
go konsek...	, і — як і було	насправді [Rp]	— це про
	прагнули б політичного визволення ...	за [Q] [Rp] [Spsa] [Spsi] [Vmen]	все , вони
coraz wielki...	європейських державах , політично ...	чимраз [Rp]	більші су
o utworzeni...	, що деякі з них —	за [Q] [Rp] [Spsa] [Spsi] [Vmen]	допомог
acy	до певної міри слушно — вважали	за [Q] [Rp] [Spsa] [Spsi] [Vmen]	панівний
czą , co wy...	. Поляки становили б економічну (	разом [Ncmsin] [Rp]	з частинс
	сама доля . Досить згадати Австро...	наприкінці [Rp]	Першої с
	Інша річ — доля тих народів	після [Rp] [Spsg]	розпаду .
	з них . Це залежало б	передусім [Rp]	від міжна
ty trzema o...	, Литвою та Україною , а	також [Rp]	між трьо
ów przebieg...	так само , як це сталося	насправді [Rp]	, якби По
prniowo asy...	не сформуватися як нації , а	просто [Rp]	поступов
wspólna w...	культурна та конфесійна спільність ...	за [Q] [Rp] [Spsa] [Spsi] [Vmen]	винятком
odzenie nar...	культурна та конфесійна спільність ...	за [Q] [Rp] [Spsa] [Spsi] [Vmen]	винятком
nym - stan...	б ( як це й було	насправді [Rp]	) експорт
ch , niechr...	Захід . Незалежно від того ,	за [Q] [Rp] [Spsa] [Spsi] [Vmen]	яких обст

Wide Context | **Meta-Information**

Name



# Using ParaConc

Parallel Concordance - [chociaż]

... iwał się maszynami ani ich nie budował, chociaż w pewnych okolicznościach wydawał się d ...  
... " > To jakiś automat – pomyślałem – chociaż żadnego nie było przedtem w pokoju, a n ...  
... m skończyć z domysłami i poznać prawdę, chociaż nie wyobrażałem sobie, jak będę ją mógł ...  
... Rzucę w nią czymś – pomyślałem – ale chociaż to był tylko sen, nie mogłem się jakoś ...  
... eg> <seg id="1225"> <s id="375:3"> Chociaż się tego spodziewałem (bo wciąż całkiem ...  
... zy...?" – a ja udałem, że nie rozumiem, chociaż doskonale rozumiałem, ale miałem ją za ...  
... órym ją opasywałem, zmienił się nagle – chociaż zewnątrz nic się nie zmieniło – stał ...  
... eg> <seg id="1347"> <s id="438:5"> Chociaż wcale już nie wierzyłem, że to ona, za ...  
... łem to za oznaki narastającej senności, chociaż Bogiem a prawdą w głębi duszy wątpiłem, ...  
... g> <seg id="1436"> <s id="474:15"> Chociaż w chwili startu zamknąłem odruchowo ocz ...  
... z znajdowała się nieco powyżej mojej, a chociaż siedziałem w kabinie, to jednak musiało ...  
... , że krzyczałem coś i wywoływałem Bazę, chociaż wiedziałem, że nie ma łączności. </s> ...  
... isja uzna, że to, co powiedziałem, jest chociaż w jednym procencie prawdopodobne, tak ż ...  
... k bez jakiegokolwiek naukowego stopnia, chociaż niejeden badacz mógłby pozazdrościć tem ...  
... uły gry, przystosujemy się i do nowych, chociaż przez jakiś czas będziemy brykać, ciska ...  
... ardziej izolowane od wszystkich innych, chociaż, naturalnie, żaden taki ślad nie może b ...  
... elkiego prawdopodobieństwa autentyczne, chociaż na ogół niesprawdzalne, ocean bowiem rz ...  
... ych olbrzymów nie tkwi w ich wyglądzie, chociaż naprawdę może przyprawić o koszarne sn ...  
... wa się często „momentalnymi maszynami”, chociaż twory te nie przypominają wcale maszyn ...  
... ha i osiemnastoma pozostałymi w niej a chociaż nigdy nie przypominają oficjalnie, że to s

<s id="133:2"> океан не користувався машинами й не будував їх, хоч за певних обставин, можливо, бу  
<s id="140:5"> «Мабуть, якийсь робот», – подумав я, хоч досі не бачив у кімнаті жодного робота, а  
<s id="338:2"> я хотів нарешті покінчити з усіма здогадками й дізнатись правду, хоча навіть не уяв  
<s id="373:20"> «Треба чимось у неї пожбурити», – подумав я, та, хоч це був тільки сон, усе-таки н  
<s id="398:2"> Хоч я й сподівався його побачити (бо все ще мимохить дошукувався бодай якоїсь логік  
<s id="398:10"> Однак ніхто не знав, що я сказав їй за п'ять днів до того. </s> <s id="398:11">  
<s id="454:4"> І хоч зовні ніщо не змінилося, – я все ще обіймав Гері, дивлячись їй у вічі, – прот  
<s id="460:5"> Хоч я вже не вірив у те, що це Гері, кожного разу, коли я впізнавав її звички, мені  
<s id="462:1"> І я подумав, що її таки почав змагати сон, хоч, відверто кажучи, в глибині душі сум  
<s id="499:8"> хоч у момент startу ракети я мимоволі заплющив очі, реактивне полум'я все ж таки ос  
<s id="690:6"> я точно пам'ятаю, що, коли вдарився шасі об хвилю, її обличчя було трохи вище від м  
<s id="710:9"> здається, я щось кричав і викликав базу, хоча знав, що зв'язку немає. </s>  
<s id="723:3"> якщо комісія визнає, що в моєму повідомленні є хоч один процент правди, і вирішить  
<s id="744:5"> На ухвалу Ради, очевидно, вплинуло те, що це відкриття зробила людина без будь-яког  
<s id="753:13"> якщо ми не зможемо спекатись «гостей», то звикнемо до них і житимемо з ними, а якщ  
<s id="923:3"> вибрано найстійкіші відбитки в пам'яті, найбільш ізольовані від усіх інших, хоча, п  
<s id="1008:4"> неправдоподібні описи форм, які створює плазма, все-таки досить точні, хоч здебіль

26 matches Polish - Original text order Strings matching: chociaż

# Current and future work

- More effort is focused on building on the corpus flexibility and compatibility with standards and other projects, adjusting to existing searching software (currently co-operating with InterCorp and PARASOL projects)

But also:

- expanding the size and variability of the corpus
- enhancing the quality of the mark-up

Planning:

- Morphosyntactic disambiguation (current priority)
- Semantic annotation (time and verb classes, feature-based vs synsets)