



Using CLARIN for Digital Research

DH2014 tutorial Workshop

Lausanne, Monday 7 July 2014

Christoph Draxler

draxler@phonetik.uni-muenchen.de

Thorsten Trippel

thorsten.trippel@uni-tuebingen.de

Dieter van Uytvanck

dieter@clarin.eu

Martin Wynne

martin@clarin.eu



WHAT IS CLARIN?

- European Research Infrastructure Consortium (**ERIC**)
- **Node** in the ESFRI roadmap
- Set of national **infrastructures**
- **Network** of technical centres, data providers, researchers, developers
- **Community** of researchers using language resources and tools to do new forms of digital research in the humanities and social sciences

WHO?

- Austria, Bulgaria, Czech Republic, Denmark, Estonia, Germany, The Netherlands, Norway, Poland, Dutch Language Union.
- Croatia, Finland, France, Italy & Sweden in negotiations to join
- Contacts and activities in almost all other European countries

WHEN?

- Preparation: 2008-2011
- Construction: 2011-2015
- Operation: 2013-
- Long-term perspective

HOW MUCH?

- Preparatory phase c. €4m from FP7
- Varying national infrastructure investments plus €1m p.a. for central coordination + related projects

WHAT DOES CLARIN DO?

- Providing language data and tools as services
- Linking repositories
- Linking tools and data
- Training, advice, support
- Promoting the use of digital text and speech among scholars across the humanities and social sciences

HOW CAN I USE CLARIN?

- Find resources
- Find tools
- Combine data and tools
- Lots more!
-





The first part of
the jigsaw –
resource
discovery
across multiple
repositories...



Virtual Language Observatory

Explore the world of language resources and technology from different perspectives










[Link](#) | [Report](#) | [Help](#)

VLO > [Faceted search](#) > [Selections](#): Old English ✕ Oxford Text Archive ✕

SEARCH

SEARCH RESULTS

20 results Showing 1 to 10

<< < 1 2 >>

<p>Alfred's Orosius [Electronic resource] Expand</p> <p>Resources: 1 other </p>
<p>Ancrene Wisse Preface [Electronic resource] Expand</p> <p>Ancrene Wisse, a Middle English 'rule' or guide for female recluses, was composed in the West Midlands in the early thirteenth century. It is a carefully-constructed work, divided into the Preface edited here and eight parts (called by the author distinciones). This edition offers a corrected form of the text of the P...</p> <p>Resources: 1 other </p>
<p>Anglo Saxon Chronicle : selections / compiled by Anita Dowsing Expand</p> <p>Resources: 1 other </p>
<p>Anglo-Saxon Charter Boundaries material [Electronic resource] Expand</p> <p>The resource consists of files made at different times and on different machines during the time of compilation. It is useful from an historical perspective only since the material is now being set up as the AHRB-funded 'Language of Landscape: Reading the Anglo-Saxon Countryside' project (started in October 2004), whic...</p>

NARROW DOWN

Use the categories below to limit the search results to those matching the selected value(s).

- LANGUAGE
- Old English** ✕
- + COLLECTION
- Oxford Text Archive** ✕
- + RESOURCE TYPE
- + SUBJECT
- FORMAT

unknown type (20)

- + ORGANISATION
- + DATA PROVIDER

[Expand all](#) +
[Collapse all](#) -
[Deselect all](#) ✕

+ SEARCH OPTIONS



...and from separate online interfaces to resources...

Your query "spoof" returned 62 hits in 44 different texts (98,313,429 words [4,048 texts]; frequency: 0.63 instances per million words)

No	Filename	Hits 1 to 50	Page 1 / 2
1	A2D 38	with his conservative values,' he said. A series of	spoof or obsessively realistic 1960s westerns (Cat Balou, Butch Cassidy and
2	A7L 1262	's class structure, but seemed all at sea with his Bond	spoof , Modesty Blaise (1966). There simply isn't any
3	A9T 40	' choice What's going on. By Stephanie Nettell FAST-MOVING comic-strip	spoof on US life and politics, Phoenix Beam's cult American Eagle
4	AAH 199	Harry Enfield. Having worked his way up via his own live	spoof chat show Vic's Big Night Out, he has latterly been
5	ABW 322	. Later, there was frenzied activity in the creation of a	spoof campaign for Tunnel Vision, a 'new' TV channel.
6	ABW 389	with loose shoes. Why travel thousands of miles to visit a	spoof Paris? It was inevitable, she supposed, when America was
7	ACP 126	the street, this club flyer advertises Newcastle's Rockshots. A	spoof on washing powder packets, the colourful T-shirts are available from the
8	ACP 2313	small screen for the last two years (the South Bank Show	spoof , Norbert Smith , was a revamp of an old idea)
9	ADR 450	factor'. Some declared Minogue-free zones and preferred to play a	spoof record instead, 'I Should Be So Yucky'. The
10	AE6 1463	Thackeray began, but does not appear to have finished, a	spoof , now in the New York Public Library, which begins:
11	AN0 351	the 1955 Berlin Tunnel (see Chapter 2), proceeded to	spoof Western intelligence for several years with false information. But because so
12	AP0 902	. Polanski arrived in Hollywood from Europe on the back of a	spoof horror film called The Fearless Vampire Killers , a parody of the
13	APU 1490	Lorna made a comedy western, a beach musical and a spy	spoof in slow succession. Hitchcock wanted her for The Birds but she
14	B3J 2977	leave here, so does Fred. He thinks he can play	spoof , but I usually come out of the pub legless with as
15	B7J 2129	I love about New Scientist is the way you print the odd	spoof article or two in the issue nearest 1 April. It does
16	B7J 2130	. It does rather spoil the fun, though, when the	spoof is too obvious — the one in the 31 March issue was
17	BNT 109	had rejected it. Funnily enough, Punch ran exactly the same	spoof , with the same book, on 10 April 1991. If
18	BP4 523	the time word got out that the whole thing had been a	spoof , there was a sizeable demand for the book and, rather
19	CBC 10844	Wind — these are just some of the hilarious works of the	spoof heavy metal band whose new book Inside Spinal Tap will fill many
20	CBC 11760	22) Mrs Robinson Lemonheads Diana's got it taped as royal	spoof hits street By Tim Satchell THE Princess of Wales has obtained a
21	CBC 11762	obtained a copy of the video version of Pallas 2, the	spoof soap opera which pokes fun at the Royal Family and which will
22	CBE 1464	Stallone, 46, says he is 'flattered' that a	spoof of his Rambo films is being made. Charlie Sheen, who
23	CBE 1465	Shots 2, said: 'I'd be flattered if a	spoof was made of one of my movies.' Travolta stays quiet
24	CBE 2549	the Conference on European Security and said his speech had been a	spoof . He wanted to show how bad things could become if the
25	CBE 8446	down bats on a Charles and Diana split. Drama at TV	spoof SATIRICAL film producer Geoff Atkinson has been forced to hurriedly re-edit his

...to cross-searching distributed resources...



Zukunft DWDS Standardsicht

DWDS-Wörterbuch

in Zukunft soll die Produktion von synthetischen Fasern gesteigert werden

etw. hat Zukunft etw. lässt eine günstige Entwicklung erwarten, ist erfolgversprechend

diese Technologie hat (keine) Zukunft

künftiger Lebensweg eines Menschen

wie stellst du dir deine Zukunft vor?

Sprachwissenschaft Tempus des zukünftigen Geschehens, Futur

in dem Satz ›ich werde ihm schreiben‹ steht ›schreiben‹ in der Zukunft

Dazu:

- Zukunftsabsicht, Zukunftsaussicht, Zukunftsbild, Zukunftsforschung, Zukunftsgemälde, Zukunftsglaube, Zukunftsmusik, Zukunftsmöglichkeit, Zukunftsperspektive, Zukunftsplan, Zukunftsproblem, Zukunftsprognose, Zukunftsprojekt, Zukunftsroman, Zukunftssicherheit, Zukunftsraum, Zukunftsvision, zukunftsfreudig, zukunftsfröh, zukunftsgläubig, zukunftsreich, zukunftsicher, zukunftsfruchtig
- zukünftig

Version: 0.4.23 | Quelle: WDG | Artikeltyp: Vollartikel

Etymologisches Wörterbuch (nach Pfeifer)

kommen, ferner: Kunft, Abkunft, Ankunft, Auskunft, Auskunftei, Einkünfte, Herkunft, Niederkunft, Übereinkunft, Unterkunft, Zukunft, zukünftig, Zusammenkunft, abkommen, Abkommen, Abkomme, abkömmlich, ankommen, Ankömmling, aufkommen, auskommen, Auskommen, auskömmlich, einkommen, Einkommen, entkommen, herkommen, Herkommen, herkömmlich, nachkommen, Nachkomme, Nachkommenschaft, Nachkömmling, niederkommen, übereinkommen, überkommen, überkommen, umkommen, unterkommen, verkommen, Verkommenheit, vorkommen, Vorkommnis, zukommen

kommen Vb. 'sich einem Ort nähern, sich von einem Ort hierher begeben, einen Ort erreichen, eintreffen, erscheinen', ahd. *qeman* (8. Jh.), mhd. (mit Verschmelzung von u und e zu o) *komen*, asächs. *kuman*, mnd. *kömen*, kämen, *kommen*, mnl. *cōmen*, nl. *komen*, aengl. *cuman*, engl. *to come*, got. *qīman*, anord. *koma*, schwed. *komma* führt mit aind. *gāmati* 'geht', toch. A *kām-*, *ka-*, *kum-*, toch. B *kā(n)m-*, *ka-*, *śem-* 'kommen', griech. *bāinein* (*baiverv*) 'gehen', lat. *venire* 'kommen', lit. *gīr̃ti* 'zur Welt kommen, geboren werden, entstehen' auf eine Wurzel ie. **gyem-* 'gehen, kommen, zur Welt kommen, geboren werden'. Im Dt. ist der ursprüngliche Anlaut *qu-* (germ. *kw-*) im Verbaladjektiv *bequem* (s. d.) erhalten. – **Kunft** f. 'das Kommen, Ankommen' (selbständiges Substantiv bis ins 19. Jh.), ahd. *kumft* (8. Jh.),

Version: 1.0.69

OpenThesaurus

Synonymgruppen für Zukunft

Synonymgruppe: **Futur, Zukunft**

Oberbegriffe: **Tempus, Zeit, Zeitform**

Version: 2014-04-14 | Quelle: OpenThesaurus

DWDS-Wortprofil 3.0

Abfragewort: Vergleichswort:

Substantiv 36

Überblick zu 'Zukunft'

abhängt absehbarer als Gegenwart als Nostalgie
als Vergangenheit berufliche entgegenblickt
entgegengeht entgegengehen entschieden Erinnerung
ferner Fernsehen Gegenwart Gegenwart als
Gegenwart wie gesichert gestalten gestaltet
Herausforderungen Jugend Markt offen sichern ungeklärt
ungewisse ungewiß unklar unsicher
Verantwortung Vergangenheit

Deutsches Textarchiv (wöchentlich aktualisiert)

Treffer: **4823**

KWiC	Datum ↓	Datum ↑	Zufällig	Links	Rechts
1	1905				Was die Zukunft dieser logischen Disziplin
2	1903				Aber die Zukunft ist kein Traum .
3	1903				Und die Menschheit der Zukunft mag wohl den Tag segnen
4	1903				takt mit der Menschheit , mit der Zukunft halten .
5	1903				tendes Farbenband in die reale Zukunft der Menschheit , in die gro
6	1903				ein ganz fernes Lichtideal der Zukunft sein könnte , wird an den A
7	1903				Ideal , die wieder einmal zu der Zukunft sich auch die älteste Verg
8	1903				Die Zukunft lag so blau .
9	1903				n Wiederaufleben der höchsten Zukunft in der tiefsten Vergangenh
10	1903				gangenheit hingegeben um der Zukunft willen .
11	1903				nehr gleichgültig sein , ob diese Zukunft doch nur eine verkappte Vr
12	1903				Die sozialen Siege der Zukunft werden nichts Halbes sein
13	1903				die Ehe sei in den Stamen der Zukunft schon auselärscht , so ren

Kernkorpus 20

Treffer: **13444**, davon anzeigbar: 11790

KWiC	Datum ↓	Datum ↑	Zufällig	Links	Rechts
1	1999				Allerdings beginnt ja die Zukunft mit jedem Augenblick neu
2	1999				rsellschaft oder die Vision einer Zukunft namens Kommunismus au
3	1999				Lindens hätte es in Berlin keine Zukunft mehr gegeben .
4	1999				mtbetriebes gegenübersähe , in Zukunft womöglich noch mehr als f
5	1999				o was schadet , jetzt und für die Zukunft .
6	1999				» , um dann hoffnungsvoll in die Zukunft zu blicken : Wir haben Lich
7	1999				verser Anfang , der Streit für die Zukunft verhielß .
8	1999				ues Lebensgefühl mitzureißen : Zukunft war doch machbar !
9	1999				mens hat der Staatshaushalt in Zukunft nichts mehr zu tun .
10	1999				Vergangenheit bzw. der fernen Zukunft an , und so blieb es nicht a
11	1999				r Ausklang lag in unbestimmter Zukunft , und so mochte es manchr
12	1999				det , damit jeder Ehegatte seine Zukunft unabhängig von materielle
13	1900				meter zu nehmen als eine vane Zukunft

...and orchestration of distributed web services.

Browser address bar: <https://weblicht.sfs.uni-tuebingen.de/WebLicht-4/>

Navigation: Main Page | Chain 1 x +

Available Annotations for: German Plain Text

- POS Tags/Lemmas
- Morphology
- Constituent Parses
- Dependency Parses
- Named Entities**

German Named Entity Recognizer

Annotation Layers: language = de

Download as Excel sheet | Download as CSV

1 / 4 [1 - 30 / 91]

token ID	tokens	POSTags	lemmas	namedEntities
t1	Damals	ADV	damals	
t2	in	APPR	in	
t3	Odessa	NE	Odessa	B-GPE
t4	war	VAFIN	sein	
t5	es	PPER	es	
t6	sehr	ADV	sehr	
t7	kalt	ADJD	kalt	

Download TCF

Input and Chain Selection

Run Tools | Clear Results | Download chain

My Input Plain Text Damals in Odessa war es sehr kalt. Wir fahren jeden Morgen mit großen rappenden Lastwagen über das Kopfsteinpflaster zum	SFS To TCF Converter Language: German Document Type: TCF TCF Version: 0.4 Text	IMS Tokenizer Sentences Tokens	IMS TreeTagger Part of Speech: STTS Tagset Lemmas	SFS German Named Entity Named Entities: tuebadz8
--	---	---	--	--

| Done running tools.



13:00 Welcome and setting the agenda *Martin Wynne*

13:30 How can I find resources using CLARIN? *Dieter van Uytvanck*

14:00 How can CLARIN archive and curate my resources? *Christoph Draxler*

14:30 *break*

15:00 How can I use CLARIN to make more impact with my research? *Martin Wynne*

15:30 How can I integrate my tools and services with the CLARIN infrastructure?
Thorsten Trippel

16:00 How can I do research in the humanities with CLARIN? *Martin Wynne*

17:00 *end*