

UiT

THE ARCTIC
UNIVERSITY
OF NORWAY

TROLLing: Open Data for Linguists

Laura A. Janda, UiT

with a lot of help from colleagues at the Department of
Language and Linguistics and the University Library



How linguistics used to work...



Why is the past tense of *dive* sometimes *dove* and sometimes *dived*?

Linguistics and data

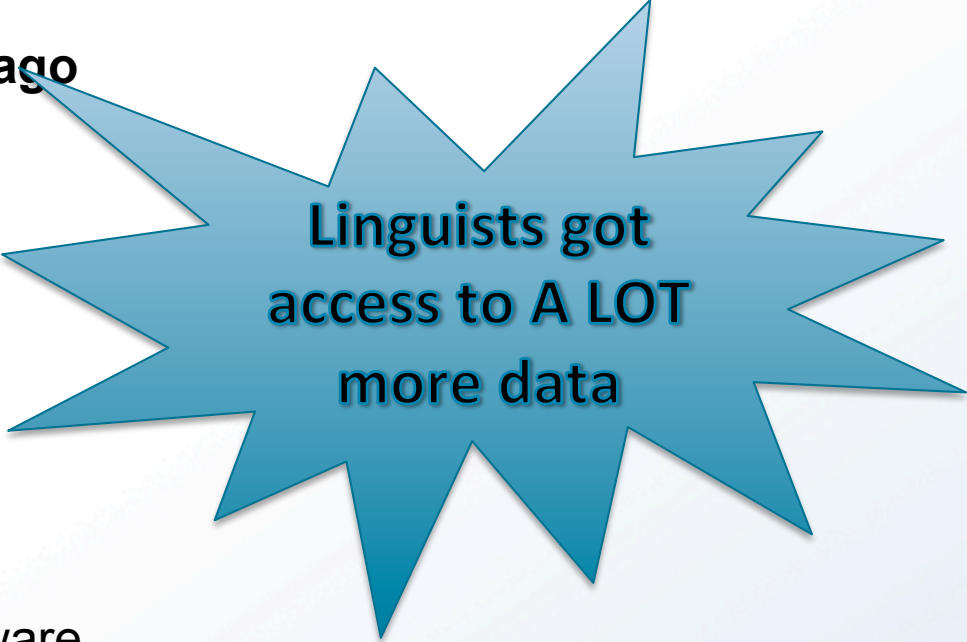
Two things happened >10 years ago

- Advent of digital corpora
 - for many languages
 - 100s of millions of words
 - balanced, annotated
- R became widely used
 - open source statistical software

Linguistics and data

Two things happened >10 years ago

- Advent of digital corpora
 - for many languages
 - 100s of millions of words
 - balanced, annotated
- R became widely used
 - open source statistical software




Linguists got
access to A LOT
more data

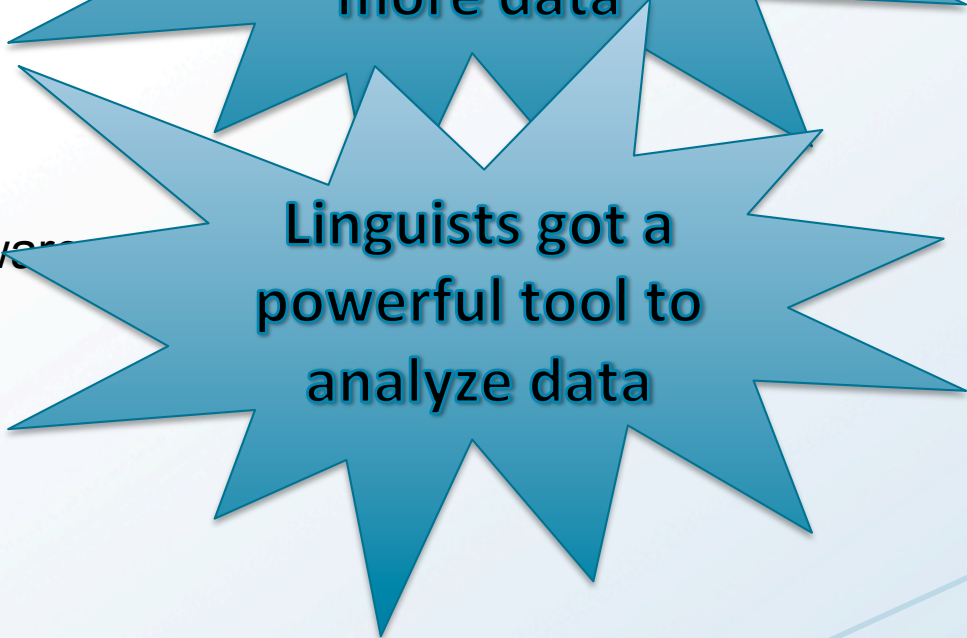
Linguistics and data

Two things happened >10 years ago

- Advent of digital corpora
 - for many languages
 - 100s of millions of words
 - balanced, annotated
- R became widely used
 - open source statistical software

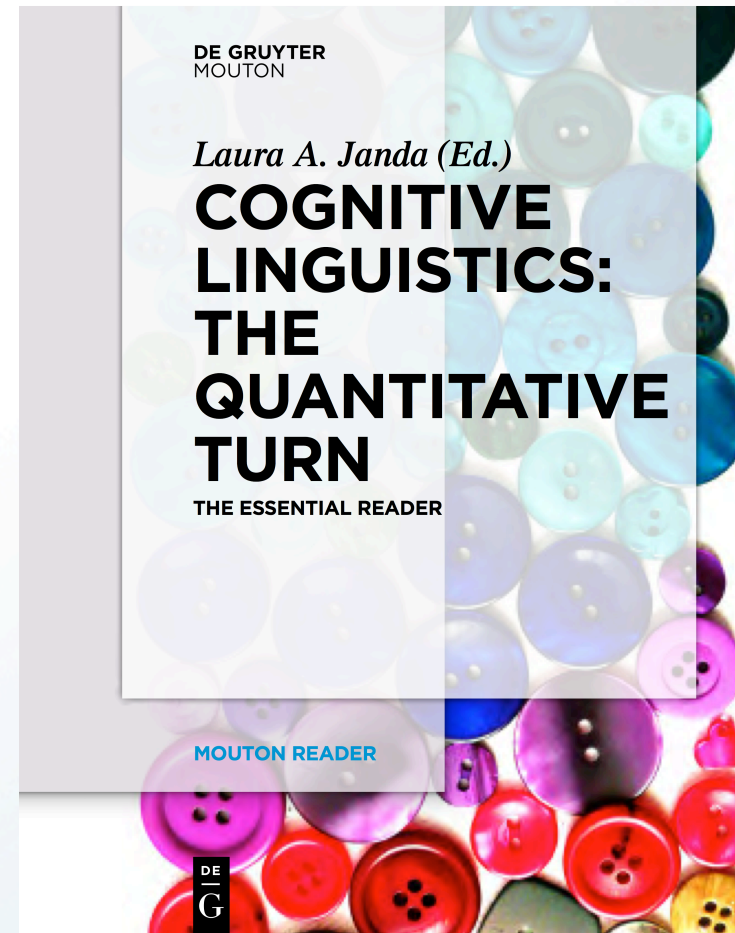
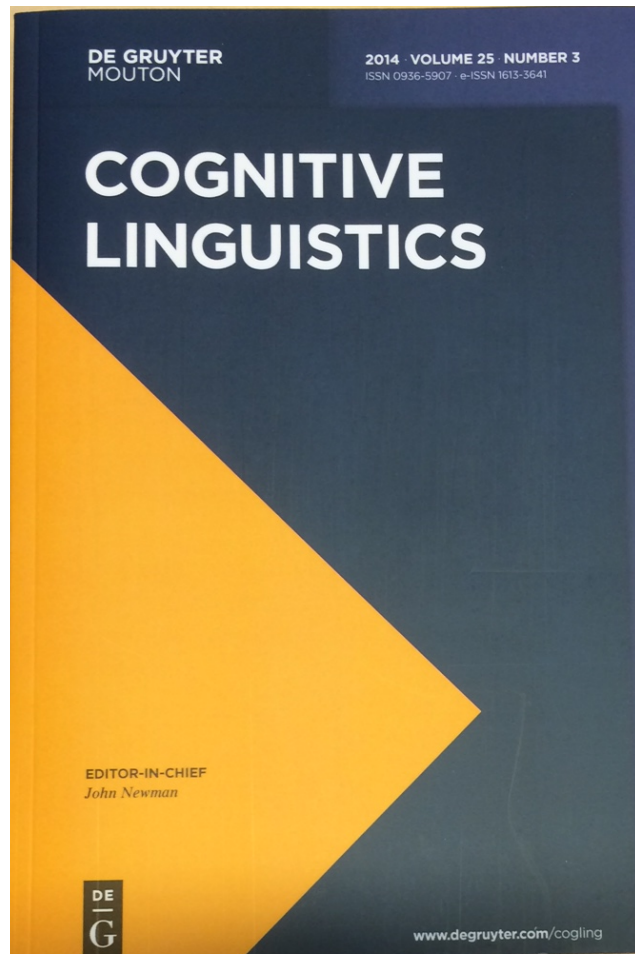


Linguists got
access to A LOT
more data

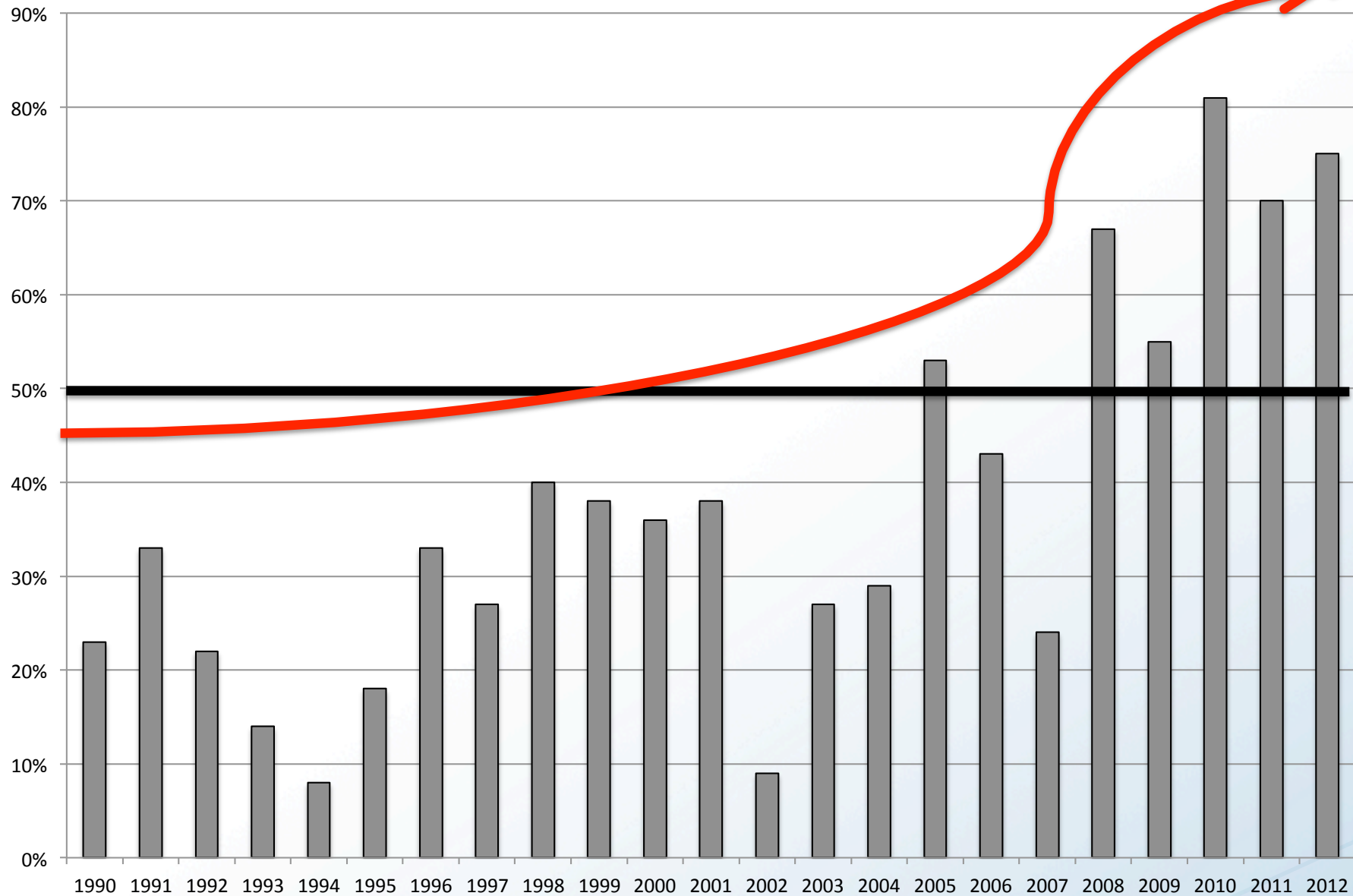


Linguists got a
powerful tool to
analyze data

The view from one journal...



percent quantitative articles in *Cognitive Linguistics* 1990-2012



How linguistics works today...



The usual process:

- Data is extracted from corpus or collected from experiments
- Laborious cleaning, tagging
- Statistical analysis
- Publications

BUT:

- What happens to the data after results are published?
- Can the researcher find and interpret the data later?
- What if someone else wants to use the data?

What today's linguists need

A PLACE TO PUBLICLY ARCHIVE DATA AND CODE

WHY? Because we need to:

- Create ethical standards for sharing of data and code
- Set norms for use of statistical methods
- Learn from each other and help our community grow
- Secure and maintain scientific integrity

What today's linguists need

A PLACE TO PUBLICLY ARCHIVE DATA AND CODE

WHY? Because we need to:

- Create ethical standards for sharing of data and code
- Set norms for use of statistical methods
- Learn from each other and help our community grow
- Secure and maintain scientific integrity

A PLACE FOR US:





TROLLing

- is an international archive of linguistic data and statistical code
- is built on the Dataverse platform from Harvard University and complies with DataCite, the international standard for storing and citing research data
- is compliant with CLARIN, the EU research infrastructure for language-based resources
- assigns a permanent URL to each post
- uses metadata that ensures visibility and retrieval through international services
- is professionally managed by the University Library of Tromsø and an international steering committee

Find TROLLing at opendata.uit.no



UiT Dataverse Network >

TROLLing Dataverse

Search   [Create Account](#) [Log In](#)


The Tromsø Repository of Language and Linguistics.






[Getting started with TROLLing](#)

TROLLing

Search Studies [Search](#)
[Advanced Search](#) [Tips](#)

Sort By:

Studies: 18 | Downloads: 177 

Replication data for: Ditransitive constructions in Russian and Ukrainian by Mykhaylyk, Roksolana; Rodina, Yulia; Anderssen, Merete Description: Publication abstract: This paper presents new experimental data on the acquisition of structures involving ditransitive verbs in two East Slavic languages: Russian and Ukrainian. The results of an elicited production task with adults and 3... Continue [+]	hdl:10037.1/10005  39 downloads Last Released: Jun 18, 2014
Replication data for: Old Church Slavonic byti Part One: Grammatical Profiling Analysis by Eckhoff, Hanne M.; Janda, Laura A.; Nessel, Tore Description: There is controversy over whether byti 'be' in Old Church Slavonic functioned as an imperfective verb with an unusually large number of inflected forms or as an aspectual pair of verbs, reflecting its suppletive origin from two stems (es- ... Continue [+]	hdl:10037.1/10008  6 downloads Last Released: May 21, 2014
Spanish deverbal adjectives (1): ivo by Fábregas, Antonio Description: Attested adjectives in -ivo in contemporary Spanish, with information about their bases, their semantic interpretation and their morphological properties.	hdl:10037.1/10018  18 downloads Last Released: Jun 16, 2014
Metonymy in Word-Formation: Russian, Czech, and Norwegian by Janda, Laura A. Description: Publication abstract: A foundational goal of cognitive linguistics is to explain linguistic phenomena in terms of general cognitive strategies rather than postulating an autonomous language module (Langacker 1987: 12-13). Metonymy is ident... Continue [+]	hdl:10037.1/10020  3 downloads Last Released: Jun 13, 2014
Why Russian Prefixes Aren't Empty by Janda, Laura A. Description: Publication abstract: We claim that Russian verbal prefixes always express meaning, even when they are used to form the perfective partners of aspectual pairs. The prefixes in verbs like написать/na-pisat' 'write' and сварить/s-varit' 'coo... Continue [+]	hdl:10037.1/10022  10 downloads Last Released: Jun 16, 2014
Russian nu-drop verbs	hdl:10037.1/10024

Getting started with TROLLing

<http://site.uit.no/trolling/getting-started/>

- Promotional video
- Instructional videos
 - User guide
- TROLLing banner