

# Restrictions on the ontological category of indefinite pronoun series in the languages of Europe

ConSOLe XXVI, UCL

Daria Bikina

[daria.bikina@gmail.com](mailto:daria.bikina@gmail.com)

HSE School of Linguistics (Moscow)

February 14-16, 2018

# Outline

## Introduction

# Outline

Introduction

Data and methods

# Outline

Introduction

Data and methods

Characteristics of pronominal series

# Outline

Introduction

Data and methods

Characteristics of pronominal series

Characteristics of pronominal systems

# Outline

Introduction

Data and methods

Characteristics of pronominal series

Characteristics of pronominal systems

Discussion

## Forming indefinite pronouns

[Haspelmath 1997]: 22

Indefinite pronouns normally occur in SERIES which have one member for each of the major ONTOLOGICAL CATEGORIES such as person, thing, property, place, time, manner, amount, plus a few others.

...

In the most common case, indefinite pronouns consist of (i) a stem indicating ontological category, plus (ii) a formal element shared by all members of indefinite pronoun series, such as *some-* and *any-* in English ...

This element will be called INDEFINITENESS MARKER.

# Forming indefinite pronouns

However, an indefinite pronoun system may not be a simple multiplication of possible stems and indefiniteness markers.



# Forming indefinite pronouns

Not just a multiplication: three Russian series

Category	koe-series	to-series	libo-series
PERSON	koe-kto	kto-to	kto-libo
THING	koe-čto	čto-to	čto-libo
PLACE.IN	koe-gde	gde-to	gde-libo
PLACE.ILL	koe-kuda	kuda-to	kuda-libo
PLACE.EL	koe-otkuda	otkuda-to	otkuda-libo
TIME	*koe-kogda	kogda-to	kogda-libo
MANNER	#koe-kak	kak-to	kak-libo
REASON	*koe-počemu	počemu-to	počemu-libo
AMOUNT	?koe-skol'ko	skol'ko-to	?skol'ko-libo
PROPERTY	#koe-kakoj	kakoj-to	kakoj-libo
CHOICE FROM A SET	koe-kotoryj	*kotoryj-to	?kotoryj-libo
POSSESSION	koe-čej	čej-to	čej-libo

# Forming indefinite pronouns

The problem of different stems: two English series

	Interrogative stem		Nominal stem	
	any-	some-	any-	some-
PERSON	*anywho	*somewho	anybody	somebody
THING	*anything	#somewhat	anything	something
PLACE	anywhere	somewhere	—	—
TIME	*anywhen	*somewhen	anytime	sometime
MANNER	anyhow	somehow	—	—
REASON	*anywhy	*somewhy	—	—

## The current research

- ▶ Do the gaps in indefinite pronoun systems occur accidentally?
- ▶ If not, what triggers the absence of a specific form in the system?
  - ▶ type of a stem?
  - ▶ properties of the ontological category?
  - ▶ semantics of indefinite pronoun series?
  - ▶ language properties?
  - ▶ areal influence?
- ▶ What theoretical implications do the results provide?

# Data

A sample of 21 European languages

- ▶ Partially based on a Haspelmath' sample ([Haspelmath 1997]: 244—317)
- ▶ This data has been extended and re-checked with native speakers to establish gaps

English

German

Dutch

Swedish

French

Italian

Modern Greek

Lithuanian

Russian

Ukrainian

Polish

Czech

Serbian

Bulgarian

Finnish

Hungarian

Moksha Mordvin

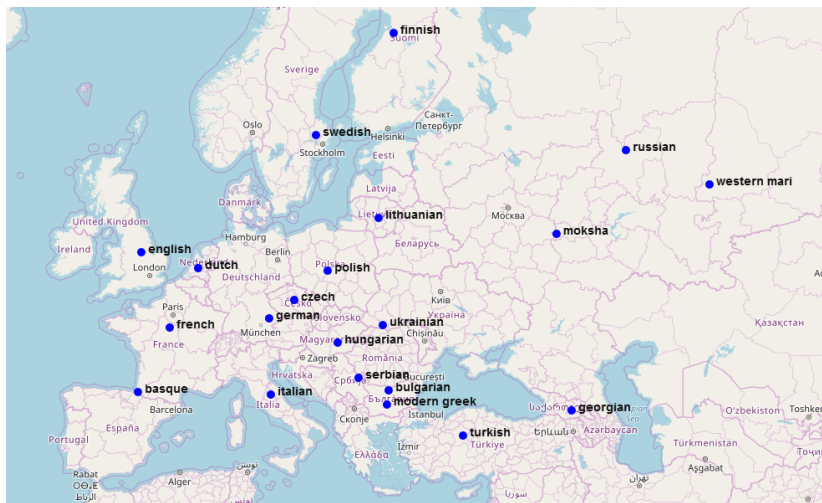
Western Mari

Basque

Turkish

Georgian

# Language sample



## Ontological categories

- ▶ The specific set of ontological categories is under discussion ([Jackendoff 1990], [Haspelmath 1997], [Award 2001], [Hengeveld, Mackenzie 2008], [Hengeveld, Mackenzie 2008])

## Ontological categories

- ▶ The specific set of ontological categories is under discussion ([Jackendoff 1990], [Haspelmath 1997], [Award 2001], [Hengeveld, Mackenzie 2008], [Hengeveld, Mackenzie 2008])
- ▶ I use the include-all-you-can approach: extended set

# Ontological categories

- ▶ The specific set of ontological categories is under discussion ([Jackendoff 1990], [Haspelmath 1997], [Award 2001], [Hengeveld, Mackenzie 2008], [Hengeveld, Mackenzie 2008])
- ▶ I use the include-all-you-can approach: extended set
  - ▶ A category is distinguished if there is a language in the sample which has a specific interrogative form of this category.



# Ontological categories

- ▶ The specific set of ontological categories is under discussion ([Jackendoff 1990], [Haspelmath 1997], [Award 2001], [Hengeveld, Mackenzie 2008], [Hengeveld, Mackenzie 2008])
- ▶ I use the include-all-you-can approach: extended set
  - ▶ A category is distinguished if there is a language in the sample which has a specific interrogative form of this category.
  - ▶ PLACE: LOCATION, PLACE: DIRECTION and PLACE: SOURCE are distinguished in many Slavic languages, including Russian: *gde* 'where', *kuda* 'to where', *otkuda* 'where from'

# Ontological categories

- ▶ The specific set of ontological categories is under discussion ([Jackendoff 1990], [Haspelmath 1997], [Award 2001], [Hengeveld, Mackenzie 2008], [Hengeveld, Mackenzie 2008])
- ▶ I use the include-all-you-can approach: extended set
  - ▶ A category is distinguished if there is a language in the sample which has a specific interrogative form of this category.
  - ▶ PLACE: LOCATION, PLACE: DIRECTION and PLACE: SOURCE are distinguished in many Slavic languages, including Russian: *gde* 'where', *kuda* 'to where', *otkuda* 'where from'
  - ▶ PROPERTY vs. CHOICE FROM A SET: English *what kind* vs. *which*

# Ontological categories

- ▶ The specific set of ontological categories is under discussion ([Jackendoff 1990], [Haspelmath 1997], [Award 2001], [Hengeveld, Mackenzie 2008], [Hengeveld, Mackenzie 2008])
- ▶ I use the include-all-you-can approach: extended set
  - ▶ A category is distinguished if there is a language in the sample which has a specific interrogative form of this category.
  - ▶ PLACE: LOCATION, PLACE: DIRECTION and PLACE: SOURCE are distinguished in many Slavic languages, including Russian: *gde* 'where', *kuda* 'to where', *otkuda* 'where from'
  - ▶ PROPERTY vs. CHOICE FROM A SET: English *what kind* vs. *which*
  - ▶ POSSESSOR: Polish *kto* 'who' vs. *czyj* 'whose'

# Ontological categories

- ▶ The specific set of ontological categories is under discussion ([Jackendoff 1990], [Haspelmath 1997], [Award 2001], [Hengeveld, Mackenzie 2008], [Hengeveld, Mackenzie 2008])
- ▶ I use the include-all-you-can approach: extended set
  - ▶ A category is distinguished if there is a language in the sample which has a specific interrogative form of this category.
  - ▶ PLACE: LOCATION, PLACE: DIRECTION and PLACE: SOURCE are distinguished in many Slavic languages, including Russian: *gde* 'where', *kuda* 'to where', *otkuda* 'where from'
  - ▶ PROPERTY vs. CHOICE FROM A SET: English *what kind* vs. *which*
  - ▶ POSSESSOR: Polish *kto* 'who' vs. *czyj* 'whose'
  - ▶ VERB for a pronominal verb: Moksha Mordvin *mej-t'-əms* (what-VBZ-INF) 'to do something')

# Ontological categories

- ▶ The specific set of ontological categories is under discussion ([Jackendoff 1990], [Haspelmath 1997], [Award 2001], [Hengeveld, Mackenzie 2008], [Hengeveld, Mackenzie 2008])
- ▶ I use the include-all-you-can approach: extended set
  - ▶ A category is distinguished if there is a language in the sample which has a specific interrogative form of this category.
  - ▶ PLACE: LOCATION, PLACE: DIRECTION and PLACE: SOURCE are distinguished in many Slavic languages, including Russian: *gde* 'where', *kuda* 'to where', *otkuda* 'where from'
  - ▶ PROPERTY vs. CHOICE FROM A SET: English *what kind* vs. *which*
  - ▶ POSSESSOR: Polish *kto* 'who' vs. *czyj* 'whose'
  - ▶ VERB for a pronominal verb: Moksha Mordvin *mej-t'-əms* (what-VBZ-INF) 'to do something')

# Ontological categories

PERSON	who
THING	what
PLACE: LOCATION	where
PLACE: DIRECTION	to where
PLACE: SOURCE	where from
TIME	when
MANNER	how
AMOUNT	how many
REASON	why
PROPERTY	what (kind)
CHOICE FROM A SET	which
POSSESSOR	whose
(VERB	to do what)

# Factors discussed

- ▶ Characteristics of **pronoun series**

# Factors discussed

- ▶ Characteristics of **pronoun series**



# Factors discussed

- ▶ Characteristics of **pronoun series**
  - ▶ Type of the stem

# Factors discussed

- ▶ Characteristics of **pronoun series**
  - ▶ Type of the stem
  - ▶ Semantics of the series

# Factors discussed

- ▶ Characteristics of **pronoun series**
  - ▶ Type of the stem
  - ▶ Semantics of the series
- ▶ Characteristics of **pronominal systems** of the sample

# Factors discussed

- ▶ Characteristics of **pronoun series**
  - ▶ Type of the stem
  - ▶ Semantics of the series
- ▶ Characteristics of **pronominal systems** of the sample
  - ▶ Absence of a non-derived (simple) interrogative of a particular category

# Factors discussed

- ▶ Characteristics of **pronoun series**
  - ▶ Type of the stem
  - ▶ Semantics of the series
- ▶ Characteristics of **pronominal systems** of the sample
  - ▶ Absence of a non-derived (simple) interrogative of a particular category
  - ▶ Number of different stem types

# Factors discussed

- ▶ Characteristics of **pronoun series**
  - ▶ Type of the stem
  - ▶ Semantics of the series
- ▶ Characteristics of **pronominal systems** of the sample
  - ▶ Absence of a non-derived (simple) interrogative of a particular category
  - ▶ Number of different stem types

## Type of the stem

- ▶ Interrogative / relative pronoun  
e.g., German irgend**wer**, irgend**was**

## Type of the stem

- ▶ Interrogative / relative pronoun  
e.g., German irgend**wer**, irgend**was**
- ▶ Another indefinite pronoun  
e.g., German jemand → irgend**jemand**;  
Hill Mari ma-gən'ät (what-INDF1) → ta-ma-gən'ät  
(INDF2-what-INDF1)



## Type of the stem

- ▶ Interrogative / relative pronoun  
e.g., German irgend**wer**, irgend**was**
- ▶ Another indefinite pronoun  
e.g., German jemand → irgend**jemand**;  
Hill Mari ma-gən'ät (what-INDF1) → ta-ma-gən'ät  
(INDF2-what-INDF1)
- ▶ Generic noun  
e. g., English any**body**, any**time**

## Type of the stem

- ▶ Interrogative / relative pronoun  
e.g., German irgend**wer**, irgend**was**
- ▶ Another indefinite pronoun  
e.g., German jemand → irgend**jemand**;  
Hill Mari ma-gən'ät (what-INDF1) → ta-ma-gən'ät  
(INDF2-what-INDF1)
- ▶ Generic noun  
e. g., English any**body**, any**time**
- ▶ Numeral 'one'  
e. g., English any**one**, French quelqu'**un**

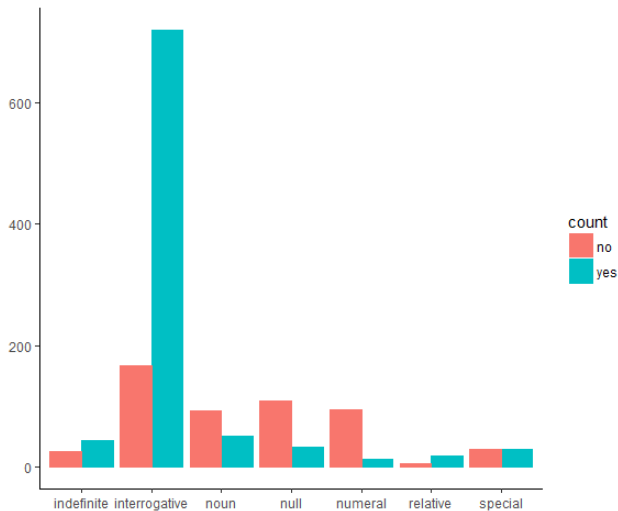
## Type of the stem

- ▶ Interrogative / relative pronoun  
e.g., German irgend**wer**, irgend**was**
- ▶ Another indefinite pronoun  
e.g., German jemand → irgend**jemand**;  
Hill Mari ma-gən'ät (what-INDF1) → ta-ma-gən'ät  
(INDF2-what-INDF1)
- ▶ Generic noun  
e. g., English any**body**, any**time**
- ▶ Numeral 'one'  
e. g., English any**one**, French quelqu'**un**
- ▶ Null stem (any, some as PROPERTY, CHOICE FROM A SET  
and AMOUNT in English)

## Type of the stem

- ▶ Interrogative / relative pronoun  
e.g., German irgend**wer**, irgend**was**
- ▶ Another indefinite pronoun  
e.g., German jemand → irgend**jemand**;  
Hill Mari ma-gən'ät (what-INDF1) → ta-ma-gən'ät  
(INDF2-what-INDF1)
- ▶ Generic noun  
e. g., English any**body**, any**time**
- ▶ Numeral 'one'  
e. g., English any**one**, French quelqu'**un**
- ▶ Null stem (any, some as PROPERTY, CHOICE FROM A SET  
and AMOUNT in English)
- ▶ Special stem type (non-regular)

# Type of the stem



## Type of the stem

Pronominal series based on numeral 'one' often are restricted to PERSON and CHOICE FROM A SET, which is probably associated with individuality

# Semantics of series

The sample of series was **very roughly** divided into 4 groups:

- ▶ Specific indefinites
- ▶ Non-specific indefinites (used in non-veridical contexts)
- ▶ Free-choice indefinites
- ▶ Negative indefinites

## Semantics of series

- ▶ The results showed that in general, semantics does not correlate with the absence of particular indefinite pronoun forms



## Semantics of series

- ▶ The results showed that in general, semantics does not correlate with the absence of particular indefinite pronoun forms
- ▶ Free-choice indefinites show fewer gaps due to the fact that many of them are sluicing-based, and thereof less grammaticalized, which implies the lack of restrictions

## Semantics of series

- ▶ The results showed that in general, semantics does not correlate with the absence of particular indefinite pronoun forms
- ▶ Free-choice indefinites show fewer gaps due to the fact that many of them are sluicing-based, and thereof less grammaticalized, which implies the lack of restrictions

Figure: Desemanticization of indefinite pronouns ([Haspelmath 1997])

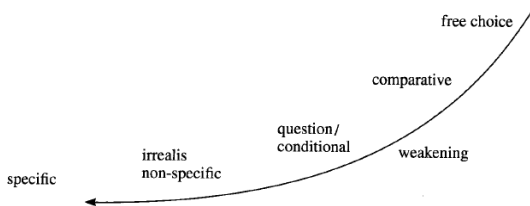
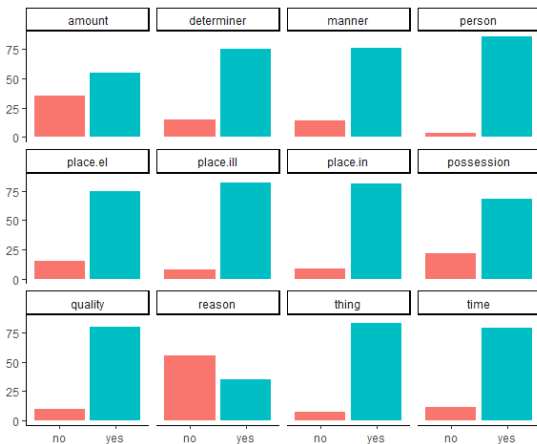


Figure: Proportion of **absent** and **present** indefinite pronoun forms by ontological category

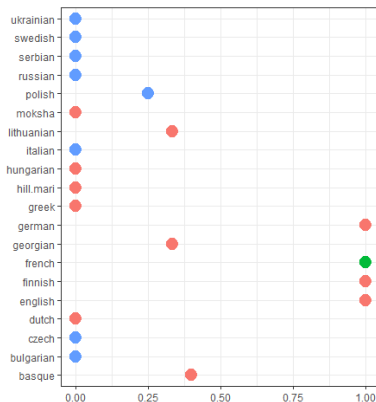


## Absence of a non-derived base interrogative

	<b>Non-derived form</b>	<b>Derived form</b>
QUANTITY	French: <i>combien</i>	English: <i>how many</i>
REASON	English: <i>why</i>	Bulgarian: <i>za-što</i> (for-what)
POSSESSOR	Russian: <i>čej</i>	Moksha: <i>kin'</i> (who.GEN)

# Absence of a non-derived base interrogative: POSSESSOR

Figure: Percentage of absent forms



Languages with special POSSESSOR pronouns; prepositional constructions; case forms of 'who' / 'which'.

# Absence of a non-derived base interrogative: POSSESSOR

POSSESSOR → PERSON

# Absence of a non-derived base interrogative: REASON

Figure: Percentage of absent forms



Languages with special REASON pronouns; prepositional constructions; dative/causal forms of 'what'.

## Absence of a non-derived base interrogative: REASON

- ▶ REASON indefinite pronouns show unexpected behaviour: not the absence, but the **presence** of a non-derived REASON interrogative triggers the absence of a corresponding indefinite form



## Absence of a non-derived base interrogative: REASON

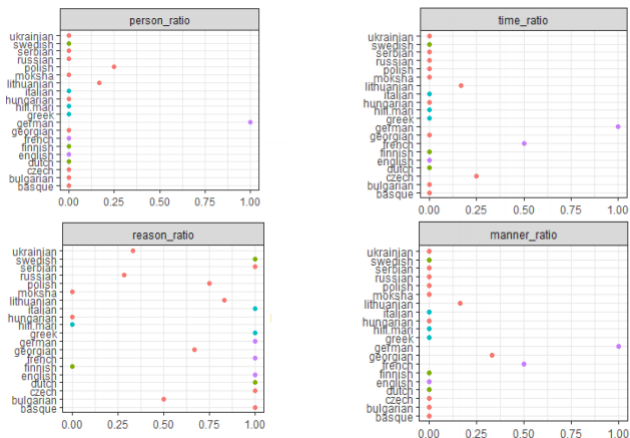
- ▶ REASON indefinite pronouns show unexpected behaviour: not the absence, but the **presence** of a non-derived REASON interrogative triggers the absence of a corresponding indefinite form
- ▶ One more argument in favour of originality of why-pronouns ([Bromberger 1992], [de Villiers 1991], [de Villiers 1996], [Rizzi 2001], [Thornton 2008])

## Absence of a non-derived base interrogative: REASON

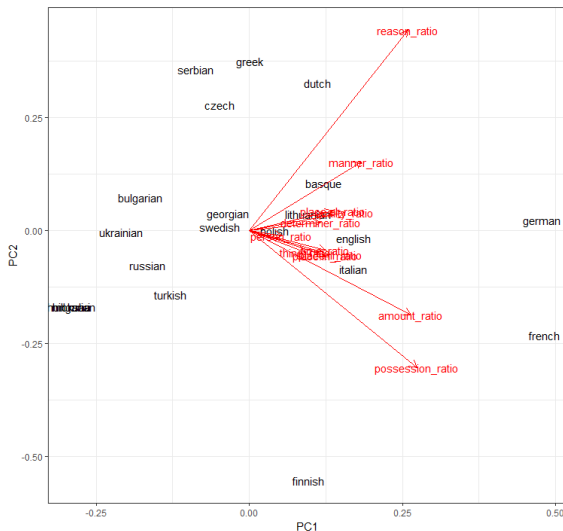
- ▶ REASON indefinite pronouns show unexpected behaviour: not the absence, but the **presence** of a non-derived REASON interrogative triggers the absence of a corresponding indefinite form
- ▶ One more argument in favour of originality of why-pronouns ([Bromberger 1992], [de Villiers 1991], [de Villiers 1996], [Rizzi 2001], [Thornton 2008])

# Number of possible stems

**Figure:** Percentage of absent forms by number of possible stems (1, 2, 3, 4) and by language: PERSON, TIME, REASON, MANNER



# PCA analysis



# Discussion

- ▶ These are the properties of languages and not the properties of particular indefinite pronoun series that influence the way a system of indefinite pronouns is organized

# Discussion

- ▶ These are the properties of languages and not the properties of particular indefinite pronoun series that influence the way a system of indefinite pronouns is organized
- ▶ Our results make a contribution to the idea of how the ontological categories may be ranked

## Discussion

- ▶ These are the properties of languages and not the properties of particular indefinite pronoun series that influence the way a system of indefinite pronouns is organized
- ▶ Our results make a contribution to the idea of how the ontological categories may be ranked [Mackenzie 2009]: There is a hierarchy of semantic categories that reflects the level of their *cognitive complexity*:

**individual**  $\subset$  **place**  $\subset$  **time**  $\subset$  **manner**  $\subset$  **quantity**  $\subset$  **reason**

- ▶ Does the percentage of gaps correlate with the frequency of corresponding stems of different ontological categories?



- ▶ Does the percentage of gaps correlate with the frequency of corresponding stems of different ontological categories?
- ▶ What exactly in the semantics of REASON pronouns conflicts with the ability to form indefinite pronouns?

# Acknowledgements






I would like to thank:

- ▶ Maria Kholodilova for supervising me,
- ▶ George Moroz for helping me with the quantitative part of the research,
- ▶ the audience of TyLex'17 at HSE for useful comments,
- ▶ and all my language consultants for their inestimable help.

This research has been supported by Russian Foundation for Basic Research, project No. 16-06-00536.

Thank you for your attention!

## References

-  J. Award. Parts of speech. *Language Typology and Language Universals, vol. 1*, 2001. Pp. 726—736.
-  S. Bromberger. On what we know we don't know. *Explanation, Theory, Linguistics, and How Questions Shape Them*. Palo Alto, CA: Stanford University, 1992.
-  J. de Villiers. Why questions? *WH: Papers in the acquisition of Wh. University of Massachusetts Occasional Papers in Linguistics*. Pp. 155—173. Amherst, MA: GLSA, 1991.
-  J. de Villiers. Defining the open and closed program for acquisition: The case of wh-questions. *Towards a genetics of language*. Pp. 145-184. Mahwah, NJ: Erlbaum, 1996.
-  M. Haspelmath. *Indefinite Pronouns*. Oxford: Clarendon Press, 1997.

## References

-  K. Hengeveld, and J. L. Mackenzie. *Functional Discourse Grammar: A typologically-based theory of language structure*. Oxford: Oxford University Press, 2008.
-  K. Hengeveld, M. L. Braga, E. de Melo Barbosa, J. S. Coriolano, J. J. da Costa, M. de Souza Martins, D. L. de Oliveira, V. M. de Oliveira, L. G. Pereira, L. Santana, C. L. do Carmo Santos, V. dos Ramos Soares. Semantic categories in the indigenous languages of Brazil. *Functions of Language* 19(1). Pp. 38—57. 2012.
-  J. L. Mackenzie. Content interrogatives in a sample of 5-languages. *Lingua* 119(8). Pp. 1131-1163. 2009.

# References

-  L. Rizzi. On the position Int (errogative) in the left periphery of the clause. *Current Studies in Italian Syntax. Essays offered to Lorenzo Renzi*. Pp. 287-296. Amsterdam: Elsevier North-Holland, 2001.
-  R. Thornton. Why continuity. *Jem Natural Language and Linguistic Theory* 26(10). Pp. 107-146. 2008.