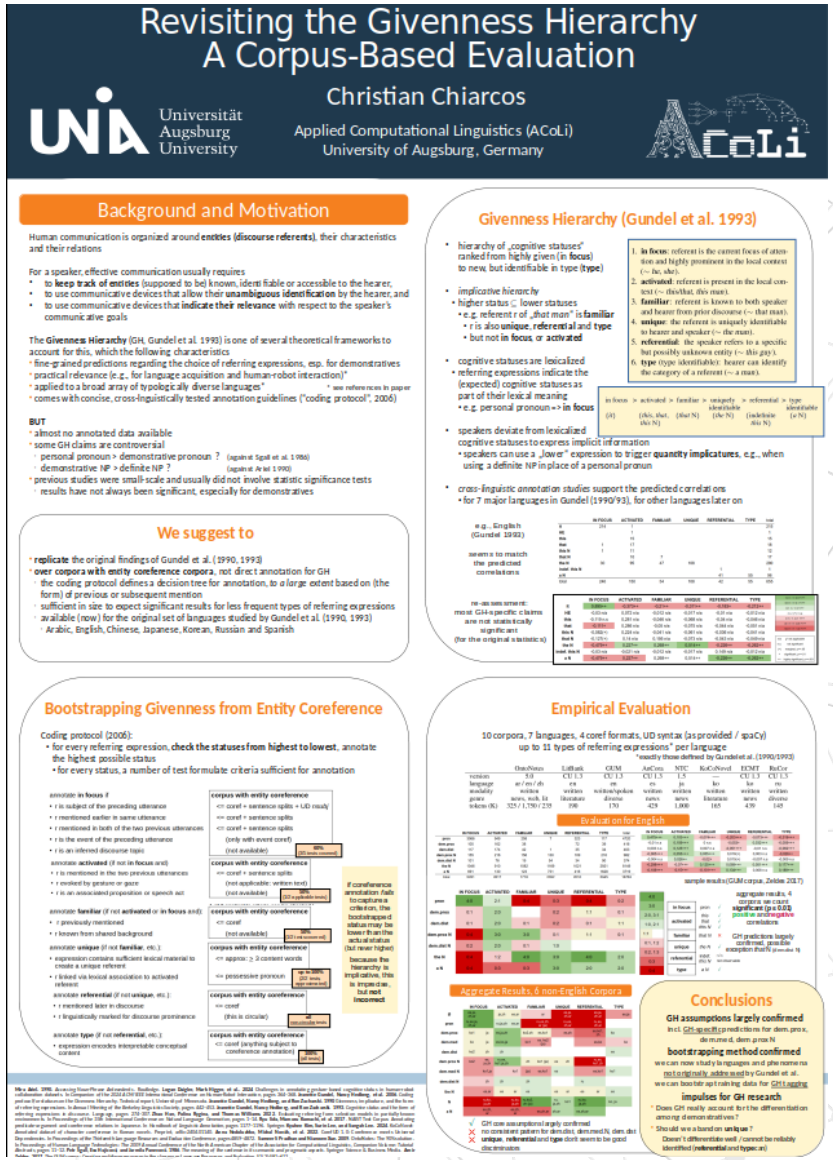


Universität Augsburg
Philologisch-Historische Fakultät

Revisiting the Givenness Hierarchy

A Corpus-Based Evaluation

Christian Chiarcos

Applied Computational Linguistics (ACoLi)

University of Augsburg, Germany

CODI/CRAC@EMNLP-2025, 2025-11-09

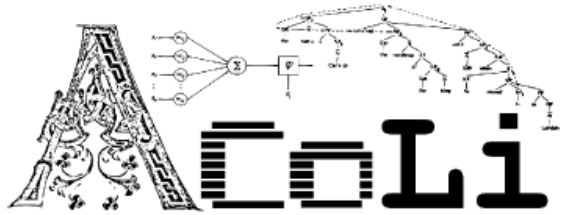
Information Status („Givenness“)

Human communication is (partially, at least) structured around **entities (discourse referents)**, their characteristics and their relations

From the perspective of a speaker, effective communication usually requires

- to **keep track of entities** (supposed to be) known, identifiable or accessible to the hearer,
- to use communicative devices that allow their **unambiguous identification** by the hearer, and
- to use communicative devices that **indicate their relevance** with respect to the speaker’s communicative goals

Givenness Hierarchy (GH, Gundel et al. 1993)



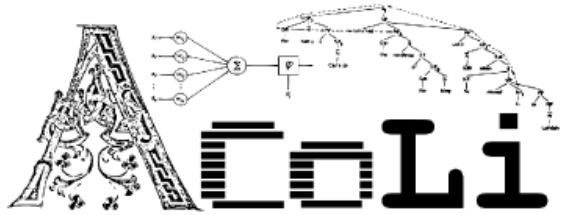
One such framework that

- makes relatively fine-grained predictions regarding the choice of referring expressions, esp. for demonstratives
- has practical relevance*

  * e.g., for language acquisition and human-robot communication, see paper
- has been applied to a broad array of typologically diverse languages**

  **although almost no data has been published, see references in paper
- comes with concise annotation guidelines („coding protocol“, 2006) applied to typologically diverse languages

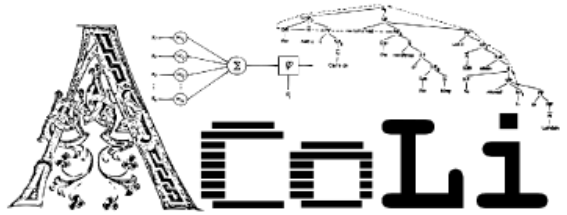
Givenness Hierarchy (GH, Gundel et al. 1993)



BUT

- almost no annotated data available
  - GH has influenced annotation efforts, but only with simplifications
- some GH claims are controversial
  - personal pronoun > demonstrative pronoun (against Sgall et al. 1986)
  - demonstrative NP > definite NP (against Ariel 1990)
- previous studies were small-scale and usually did not involve statistic significance tests
  - and their results have not always been significant, especially for demonstratives

Revisiting the Givenness Hierarchy



We suggest to

- **replicate** the original findings of Gundel et al. (1990, 1993)
- over **coreference corpora**, not direct annotation for GH (or IS)
  - the coding protocol defines a decision tree for annotation, *to a large extent* based on (the form) of previous or subsequent mention
  - sufficient in size to expect significant results for less frequent types of referring expressions
  - available (now) for the original set of languages studied by Gundel et al. (1990, 1993)

    Arabic, English, Chinese, Japanese, Korean, Russian and Spanish

Givenness Hierarchy

The Givenness Hierarchy (Gundel et al. 1993)

- *hierarchy* of „cognitive statuses“ ranked from highly given (**in focus**) to new, but identifiable in type (**type**)

1. **in focus**: referent is the current focus of attention and highly prominent in the local context (~ *he, she*).
2. **activated**: referent is present in the local context (~ *this/that, this man*).
3. **familiar**: referent is known to both speaker and hearer from prior discourse (~ *that man*).
4. **unique**: the referent is uniquely identifiable to hearer and speaker (~ *the man*).
5. **referential**: the speaker refers to a specific but possibly unknown entity (~ *this guy*).
6. **type** (type identifiable): hearer can identify the category of a referent (~ *a man*).

The Givenness Hierarchy (Gundel et al. 1993)

- *hierarchy* of „cognitive statuses“
- *implicative* hierarchy
  - higher statuses are subsets of lower statuses

    if referent *r* of „*that man*“ is **familiar**
    - *r* is also **unique**, **referential** and **type** identifiable
    - *r* is not **in focus**, or **activated**

1. **in focus**: referent is the current focus of attention and highly prominent in the local context (∼ *he, she*).
2. **activated**: referent is present in the local context (∼ *this/that, this man*).
3. **familiar**: referent is known to both speaker and hearer from prior discourse (∼ *that man*).
4. **unique**: the referent is uniquely identifiable to hearer and speaker (∼ *the man*).
5. **referential**: the speaker refers to a specific but possibly unknown entity (∼ *this guy*).
6. **type** (type identifiable): hearer can identify the category of a referent (∼ *a man*).

The Givenness Hierarchy (Gundel et al. 1993)

- *hierarchy* of „cognitive statuses“
- *implicative* hierarchy
- cognitive statuses are *lexicalized*
  - referring expressions indicate the (expected) cognitive statuses **as part of their lexical meaning**

    personal pronoun => **in focus** (in English)

1. **in focus**: referent is the current focus of attention and highly prominent in the local context (∼ *he, she*).
2. **activated**: referent is present in the local context (∼ *this/that, this man*).
3. **familiar**: referent is known to both speaker and hearer from prior discourse (∼ *that man*).
4. **unique**: the referent is uniquely identifiable to hearer and speaker (∼ *the man*).
5. **referential**: the speaker refers to a specific but possibly unknown entity (∼ *this guy*).
6. **type** (type identifiable): hearer can identify the category of a referent (∼ *a man*).

The Givenness Hierarchy (Gundel et al. 1993)

- *hierarchy* of „cognitive statuses“
- *implicative* hierarchy
- cognitive statuses are *lexicalized*
- speakers can *deviate* from the lexicalized cognitive statuses to express implicit information
  - speakers can use a „lower“ expression to trigger **quantity implicatures**, e.g., when using a definite NP in place of a personal pronun

1. **in focus**: referent is the current focus of attention and highly prominent in the local context (∼ *he, she*).
2. **activated**: referent is present in the local context (∼ *this/that, this man*).
3. **familiar**: referent is known to both speaker and hearer from prior discourse (∼ *that man*).
4. **unique**: the referent is uniquely identifiable to hearer and speaker (∼ *the man*).
5. **referential**: the speaker refers to a specific but possibly unknown entity (∼ *this guy*).
6. **type** (type identifiable): hearer can identify the category of a referent (∼ *a man*).

The Givenness Hierarchy (Gundel et al. 1993)

- *hierarchy* of „cognitive statuses“
- *implicative* hierarchy
- cognitive statuses are *lexicalized*
- speakers can *deviate* from the lexicalized cognitive statuses to express implicit meaning
- *cross-linguistic annotation studies* support the predicted correlations
  - for 7 major languages in the original publications, for many other languages later on

1. **in focus**: referent is the current focus of attention and highly prominent in the local context (∼ *he, she*).
2. **activated**: referent is present in the local context (∼ *this/that, this man*).
3. **familiar**: referent is known to both speaker and hearer from prior discourse (∼ *that man*).
4. **unique**: the referent is uniquely identifiable to hearer and speaker (∼ *the man*).
5. **referential**: the speaker refers to a specific but possibly unknown entity (∼ *this guy*).
6. **type** (type identifiable): hearer can identify the category of a referent (∼ *a man*).

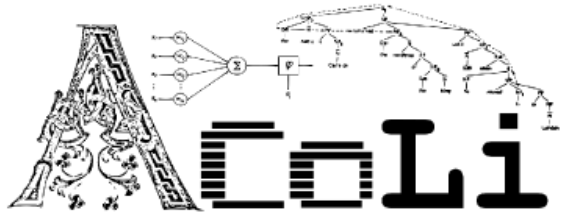
Gundel et al.‘s (1993) observations for English

	IN FOCUS	ACTIVATED	FAMILIAR	UNIQUE	REFERENTIAL	TYPE	total
it	214	1					215
HE		1					1
this		15					15
that	1	17					18
this N	1	11					12
that N		10	7				17
the N	30	95	47	108			280
indef. this N					1		1
a N					41	55	96
total	246	150	54	108	42	55	655

prediction

in focus	> activated	> familiar	> uniquely identifiable	> referential	> type identifiable
{*it*}	{*this*, *that*, *this* N}	{*that* N}	{*the* N}	{indefinite *this* N}	{*a* N}

Significance and correlation tests over Gundel et al. (1993)



	IN FOCUS	ACTIVATED	FAMILIAR	UNIQUE	REFERENTIAL	TYPE
it	0,895++	-0,373++	-0,21++	-0,311++	-0,183+	-0,212++
HE	-0,03 n/a	0,072 n/a	-0,012 n/a	-0,017 n/a	-0,01 n/a	-0,012 n/a
this	-0,119 n.s	0,281 n/a	-0,046 n/a	-0,068 n/a	-0,04 n/a	-0,046 n/a
that	-0,111+	0,286 n/a	-0,05 n/a	-0,075 n/a	-0,044 n/a	-0,051 n/a
this N	-0,082(+)	0,224 n/a	-0,041 n/a	-0,061 n/a	-0,036 n/a	-0,041 n/a
that N	-0,127(+)	0,14 n/a	0,195 n/a	-0,073 n/a	-0,043 n/a	-0,049 n/a
the N	-0,479++	0,227++	0,268++	0,514++	-0,226++	-0,262++
indef. this N	-0,03 n/a	-0,021 n/a	-0,012 n/a	-0,017 n/a	0,149 n/a	-0,012 n/a
a N	-0,479++	0,227++	0,268++	0,514++	0,226++	0,262++

1.0 > r > 0.5 +/++
0.5 > r > 0.1 +/++
0.1 > r > 0 +/++
n.s / (+) / n.a
0 > r > -0.1 +/++
-0.1 > r > -0.5 +/++
-0.5 > r > -1.0 +/++

n/a	χ^2 not applicable
n.s.	not significant
(+)	marginal, p<=.05
+	significant, p<=.01
++	highly significant, p<=.001

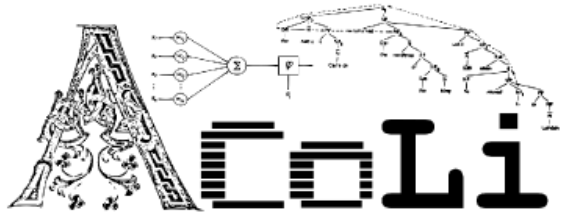
re-assessment: most GH-specific claims are not statistically significant

in focus > uniquely identifiable > type identifiable

{*it*} {*that?*} {*the* N} {*a* N}

Bootstrapping Givenness from Coreference

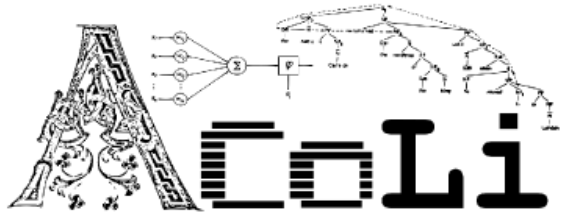
Bootstrapping Givenness from Entity Coreference



Coding protocol (2006)

- for every referring expression, check the statuses from highest to lowest, annotate the highest possible status
- annotation of status licensed by tests over context conditions (independent from surface form)
  - every test formulates a criterion sufficient for annotation
- if coreference annotation *fails* to capture a specific criterion, the bootstrapped status may be lower than the actual status
  - because the hierarchy is implicative, this is less precise, but **not incorrect**

Bootstrapping Givenness from Entity Coreference



Coding protocol (2006)

- for every referring expression, check the statuses from highest to lowest, annotate the highest possible status

annotate **in focus** if

- r is subject of the preceding utterance
- r mentioned earlier in same utterance
- r mentioned in both of the two previous utterances
- r is the event of the preceding utterance
- r is an inferred discourse topic

corpus with entity coreference

<= coref + sentence splits + UD *nsubj*

<= coref + sentence splits

<= coref + sentence splits

(only with event coref)

(not available)

3 of 5 criteria can be checked !

Bootstrapping Givenness from Entity Coreference

Coding protocol (2006)

- for every referring expression, check the statuses from highest to lowest, annotate the highest possible status

annotate **activated** (if not **in focus** and)

- r is mentioned in the two previous utterances
- r evoked by gesture or gaze
- r is an associated proposition or speech act

corpus with entity coreference

<= coref + sentence splits

(not applicable: written text)

(not available)

1 of 2 applicable criteria can be checked !

Bootstrapping Givenness from Entity Coreference

Coding protocol (2006)

- for every referring expression, check the statuses from highest to lowest, annotate the highest possible status

annotate **familiar** (if not **activated** or **in focus** and):

- r previously mentioned
- r known from shared background

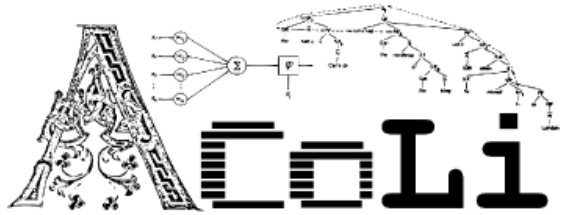
corpus with entity coreference

<= coref

(not available)

1 of 2 criteria can be checked !

Bootstrapping Givenness from Entity Coreference



Coding protocol (2006)

- for every referring expression, check the statuses from highest to lowest, annotate the highest possible status

annotate **unique** (if not **familiar**, etc.):

- expression contains sufficient lexical material to create a unique referent
- r linked via lexical association to activated referent

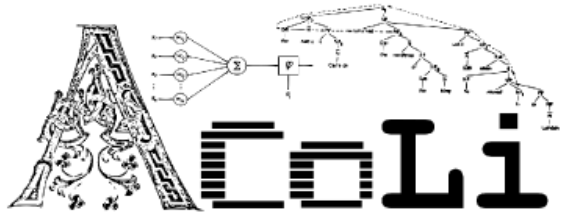
corpus with entity coreference

<= approx: ≥ 3 content words

<= possessive pronoun

all criteria can be heuristically approximated

Bootstrapping Givenness from Entity Coreference



Coding protocol (2006)

- for every referring expression, check the statuses from highest to lowest, annotate the highest possible status

annotate **referential** (if not **unique**, etc.):

- r mentioned later in discourse
- r linguistically marked for discourse prominence

corpus with entity coreference

<= coref

(this is circular)

all *context* criteria can be automatically checked

Bootstrapping Givenness from Entity Coreference

Coding protocol (2006)

- for every referring expression, check the statuses from highest to lowest, annotate the highest possible status

annotate **type** (if not **referential**, etc.):

- expression encodes interpretable conceptual content

corpus with entity coreference

<= coref (anything subject to coreference annotation)

all criteria can be automatically checked

Bootstrapping Givenness from Entity Coreference

Coding protocol (2006)

- for every referring expression, check the statuses from highest to lowest, annotate the highest possible status
- Bootstrapping
  - **in focus** (3/5 applicable tests, 60%)
  - **activated** (1/2 applicable tests, 50%)
  - **familiar** (1/2 applicable tests, 50%)
  - **unique** (2/2* applicable tests, 100%)
  - **referential** (1/1 applicable tests, 100%)
  - **type** (1/1 applicable tests, 100%)

if coreference annotation *fails* to capture a criterion, the bootstrapped status may be lower than the actual status

- because the hierarchy is implicative, this is less precise, but **not incorrect**

* approximative

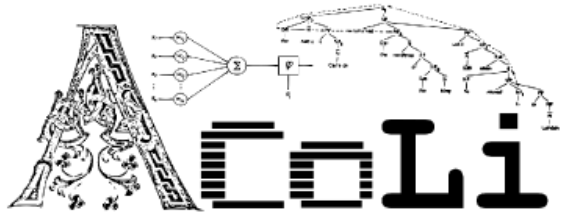
Evaluation

Data and Preprocessing

- preprocessing pipelines for 10 corpora in four formats
  - OntoNotes XML (3 corpora)
  - CorefUD (5 corpora)
  - Japanese NTC 1.5 corpus (via our CorefUD conversion)
  - Korean KoCoNovel (via our CorefUD conversion)

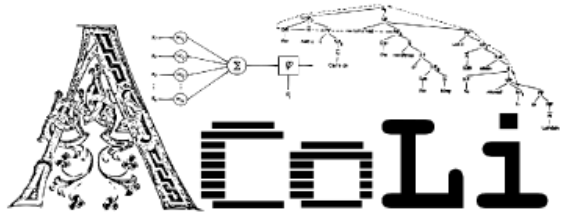
	OntoNotes	LitBank	GUM	AnCora	NTC	KoCoNovel	ECMT	RuCor
version	5.0	CU 1.3	CU 1.3	CU 1.3	1.5	—	CU 1.3	CU 1.3
language	ar / en / zh	en	en	es	ja	ko	ko	ru
modality	written	written	written/spoken	written	written	written	written	written
genre	news, web, lit	literature	diverse	news	news	literature	news	diverse
tokens (K)	325 / 1,750 / 235	190	170	429	1,000	165	439	145

Data and Preprocessing



- preprocessing pipelines for 10 corpora in four formats
- for subject detection, we use Universal Dependencies (UD) annotations
  - needed for one **in focus** criterion
  - if not provided by the source, created on-the-fly using native spaCy and (for Arabic) UDPipe models

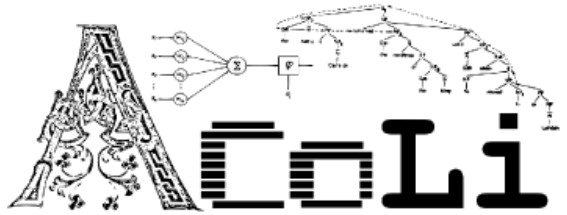
Data and Preprocessing



- preprocessing pipelines for 10 corpora in four formats
- for subject detection, we use Universal Dependencies (UD) annotations
- following Gundel et al. (1993), we distinguish up to 11 types of referring expressions per language
  - Ø (zero anaphora, only annotated in some corpora)
  - pron (personal pronoun)
  - dem.prox/med/dist (proximal/medial/distal demonstrative pronoun)
  - dem.prox N, dem.med N, dem.dist N (demonstrative NPs)
  - def N (NP with definite determiner)
  - Ø N (bare NP*)
  - a (one) N (indefinite or one-NP*)

* where considered by Gundel et al. (1993/1990)

Findings for English: GUM Corpus



	IN FOCUS	ACTIVATED	FAMILIAR	UNIQUE	REFERENTIAL	TYPE	total
pron	3066	949	258	7	323	117	4720
dem.prox	105	162	38		72	39	416
dem.dist	117	174	42	1	35	34	403
dem.prox N	155	216	156	130	109	216	982
dem.dist N	101	76	19	54	34	90	374
the N	1046	910	1082	1169	1021	2921	8149
a N	691	130	123	731	416	1628	3719
total	5281	2617	1718	2092	2010	5045	18763

total = referring expressions with coref annotation that correspond to one of the Gundel et al. (1990/93) types

	IN FOCUS	ACTIVATED	FAMILIAR	UNIQUE	REFERENTIAL	TYPE
pron	0,475+++	0,103+++	-0,074+++	-0,203+++	-0,073+++	-0,319+++
dem.prox	-0,01 n.s	0,109+++	0 n.s	-0,053+	0,032+++	-0,059+++
dem.dist	0,003 n.s	0,125+++	0,007 n.s	-0,051+++	-0,01 n.s	-0,062+++
dem.prox N	-0,065+++	0,055+++	0,055+++	0,016(+)	0,003 n.s	-0,026++
dem.dist N	-0,004 n.s	0,026++	-0,02+	0,015(+)	-0,007 n.s	-0,009 n.s
the N	-0,298+++	-0,07+++	0,125+++	0,089+++	0,051+++	0,177+++
a N	-0,106+++	-0,15+++	-0,101+++	0,134+++	0,008 n.s	0,189+++

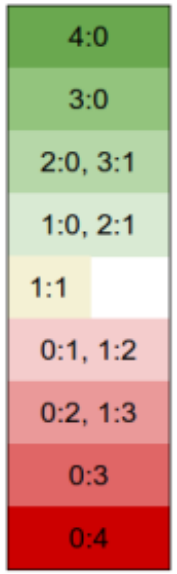
FERENTIAL	TYPE
-0,183+	-0,212++
-0,01 n/a	-0,012 n/a
-0,04 n/a	-0,046 n/a
0,044 n/a	-0,051 n/a
0,036 n/a	-0,041 n/a
0,043 n/a	-0,049 n/a
-0,226++	-0,262++
0,149 n/a	-0,012 n/a
-0,226++	-0,262++

1.0 > r > 0.5	+/++
0.5 > r > 0.1	+/++
0.1 > r > 0	+/++
n.s / (+) / n.a	
0 > r > -0.1	+/++
-0.1 > r > -0.5	+/++
-0.5 > r > -1.0	+/++

n/a	χ^2 not applicable
n.s.	not significant
(+)	marginal, p<=.05
+	significant, p<=.01
++	highly significant, p<=.001

Findings for English: Aggregate Results from 4 Corpora

	IN FOCUS	ACTIVATED	FAMILIAR	UNIQUE	REFERENTIAL	TYPE
pron	4:0	2:1	0:4	0:3	0:4	0:2
dem.prox	0:1	2:0		0:2	1:1	0:1
dem.dist	0:1	2:0	0:1	0:2	0:1	1:1
dem.prox N	0:4	3:0	3:0	0:1	1:1	0:1
dem.dist N	0:2	2:0	0:1	1:0		
the N	0:4	1:2	4:0	3:0	4:0	2:0
a N	0:4	0:3	0:3	3:0	2:0	2:0



counting
significant ($p \leq 0.01$)
positive and **negative**
correlations over 4
corpora

in focus	pron	✓
activated	*this*	✓
	that	✓
	this N	✓
familiar	*that N*	✗
unique	*the N*	✓
referential	indef.	n/a
	this N	not observable
type	*a N*	✓

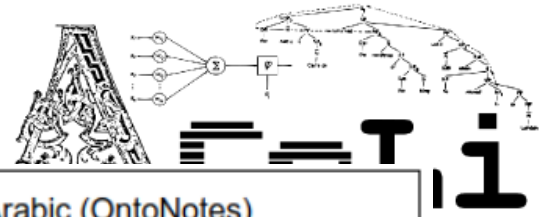
overall, the findings for English support the GH
some exceptions may be artifacts of the bootstrapping heuristics

- *this N* ~ **familiar**
- *a N* ~ **unique**
- *the N* ~ **type/referential**

In disagreement with GH, we find

- *that N* ~ **activated**

Findings for non-English: Aggregate Results from 6 Corpora



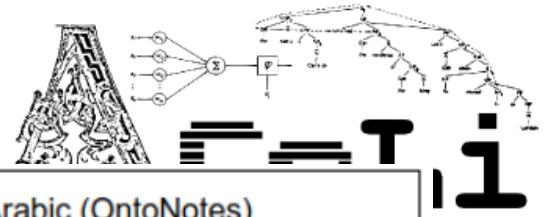
	IN FOCUS		ACTIVATED		FAMILIAR		UNIQUE		REFERENTIAL		TYPE
Ø	es,ja, zh,ar		ja,zh	es,ar		ar		es,ja, zh,ar		es,ja, zh,ar	es,ja
pron	ru,es,ja, zh,ar		ru,ja,zh	es,ar		ru,es,zh, ar (ja)		ru,es, zh,ar		ru,es, zh,ar	
dem.prox	ko1	ja	es,ja,zh		ko2,zh	es,ko1		es,zh		es,ko1, zh	ko
dem.med	ko	ja	es,ko,ja		ko1	es, ko2 (ja)				es,ko	ko
dem.dist	ko2	zh	zh								es
dem.prox N	ko2	es, ja,zh	ru,es, ko1,ja,zh		zh	ko1 (ja)	es	zh		ru,es, ko1,zh	
dem.med N		ko1,ja		ko1	(ja)	es,ko1	es			es,ko1	ko1
dem.dist N		zh	zh		zh				ru		
the N		es,ar	es	ar	es		es	ar	es	ar	es
N		ru,ko, ja,zh		ru,ko, ja,zh	ru,ko2, zh (ja)		ru, ja,zh	ko1	ru,ko, ja,zh		ko, ja
a N		es,zh, ar		es,zh, ar		es,zh,ar	zh,ar		es,zh,ar		

ar	Arabic (OntoNotes)
es	Spanish (AnCora)
ja	Japanese (NTC)
ko1	Korean (ECMT only)
ko2	Korean (KoCoNovel only)
ko	Korean (ECMT=KoCoNovel)
ru	Russian (RuCor)
zh	Chinese (OntoNotes)

4:0, 5:0
3:0
2:0, 3:1
1:0, 2:1, 3:2
1:1, 2:2
0:1, 1:2, 2:3
0:2, 1:3
0:3
0:4, 0:5

✓ GH core assumptions confirmed
Ø, pron, dem.prox, dem.med, dem.prox N, Ø N as predicted
demonstrative **type** might be event anaphor

Findings for non-English: Aggregate Results from 6 Corpora



	IN FOCUS		ACTIVATED		FAMILIAR		UNIQUE		REFERENTIAL		TYPE
Ø	es,ja, zh,ar		ja,zh	es,ar		ar		es,ja, zh,ar		es,ja, zh,ar	es,ja
pron	ru,es,ja, zh,ar		ru,ja,zh	es,ar		ru,es,zh, ar (ja)		ru,es, zh,ar		ru,es, zh,ar	
dem.prox	ko1	ja	es,ja,zh		ko2,zh	es,ko1		es,zh		es,ko1, zh	ko
dem.med	ko	ja	es,ko,ja		ko1	es, ko2 (ja)				es,ko	ko
dem.dist	ko2	zh	zh								es
dem.prox N	ko2	es, ja,zh	ru,es, ko1,ja,zh		zh	ko1 (ja)	es	zh		ru,es, ko1,zh	
dem.med N		ko1,ja		ko1	(ja)	es,ko1	es			es,ko1	ko1
dem.dist N		zh	zh		zh				ru		
the N		es,ar	es	ar	es		es	ar	es	ar	es
N		ru,ko, ja,zh		ru,ko, ja,zh	ru,ko2, zh (ja)		ru, ja,zh	ko1	ru,ko, ja,zh		ko, ja
a N		es,zh, ar		es,zh, ar		es,zh,ar	zh,ar		es,zh,ar		

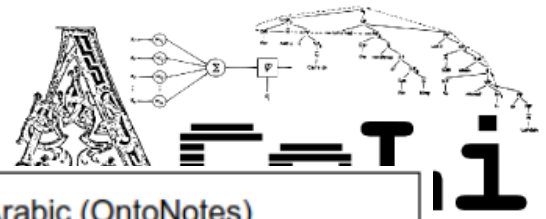
ar	Arabic (OntoNotes)
es	Spanish (AnCora)
ja	Japanese (NTC)
ko1	Korean (ECMT only)
ko2	Korean (KoCoNovel only)
ko	Korean (ECMT=KoCoNovel)
ru	Russian (RuCor)
zh	Chinese (OntoNotes)

4:0, 5:0
3:0
2:0, 3:1
1:0, 2:1, 3:2
1:1, 2:2
0:1, 1:2, 2:3
0:2, 1:3
0:3
0:4, 0:5

✓ GH core assumptions confirmed

✗ no consistent pattern for dem.dist, dem.med.N, dem.dist N
(Maybe givenness isn't the decisive factor, here?)

Findings for non-English: Aggregate Results from 6 Corpora



	IN FOCUS		ACTIVATED		FAMILIAR		UNIQUE		REFERENTIAL		TYPE
Ø	es,ja, zh,ar		ja,zh	es,ar		ar		es,ja, zh,ar		es,ja, zh,ar	es,ja
pron	ru,es,ja, zh,ar		ru,ja,zh	es,ar		ru,es,zh, ar (ja)		ru,es, zh,ar		ru,es, zh,ar	
dem.prox	ko1	ja	es,ja,zh		ko2,zh	es,ko1		es,zh		es,ko1, zh	ko
dem.med	ko	ja	es,ko,ja		ko1	es, ko2 (ja)				es,ko	ko
dem.dist	ko2	zh	zh								es
dem.prox N	ko2	es, ja,zh	ru,es, ko1,ja,zh		zh	ko1 (ja)	es	zh		ru,es, ko1,zh	
dem.med N		ko1,ja		ko1	(ja)	es,ko1	es			es,ko1	ko1
dem.dist N		zh	zh		zh				ru		
the N		es,ar	es	ar	es		es	ar	es	ar	es
N		ru,ko, ja,zh		ru,ko, ja,zh	ru,ko2, zh (ja)		ru, ja,zh	ko1	ru,ko, ja,zh		ko, ja
a N		es,zh, ar		es,zh, ar		es,zh,ar	zh,ar		es,zh,ar		

ar	Arabic (OntoNotes)
es	Spanish (AnCora)
ja	Japanese (NTC)
ko1	Korean (ECMT only)
ko2	Korean (KoCoNovel only)
ko	Korean (ECMT=KoCoNovel)
ru	Russian (RuCor)
zh	Chinese (OntoNotes)

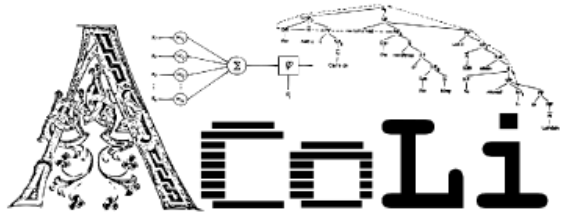
4:0, 5:0
3:0
2:0, 3:1
1:0, 2:1, 3:2
1:1, 2:2
0:1, 1:2, 2:3
0:2, 1:3
0:3
0:4, 0:5

✓ GH core assumptions confirmed

❌ no consistent pattern for dem.dist, dem.med.N, dem.dist

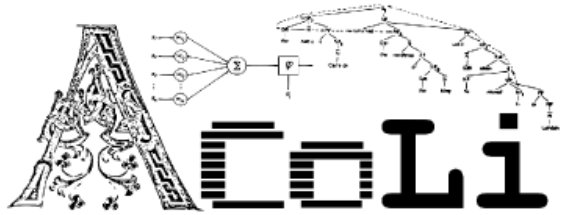
❌ **unique**, **referential** and **type** don't seem to be good discriminators

Conclusion



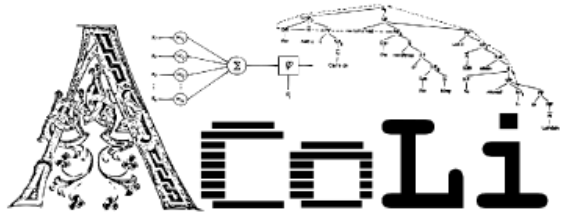
- correlations confirm GH core assumptions
  - this is also a verification of the bootstrapping method
  - possible exception are medial and distal demonstrative NPs
  - most demonstratives are licensed by **activated**, but there is little evidence that indicate that GH's cognitive statuses account for differences between demonstratives
    - maybe, these serve other communicative functions

Conclusion



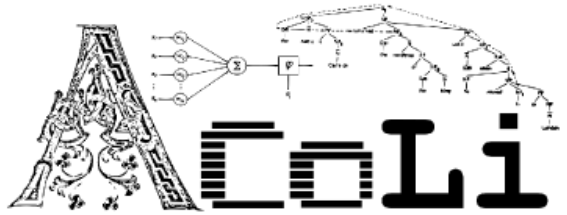
- correlations confirm GH core assumptions
- we can induce language-specific Givenness Hierarchies from coreference annotations
  - we can extend the GH to grammatical devices not addressed before (paper: case study for proper names)
  - we can extend the GH to other languages with coreference annotation
  - we can explore the role of GH in language evolution by comparing the GHs of related languages
  - as the GH has some technical relevance for human-machine interaction, this means we can support a broader set of languages

Conclusion



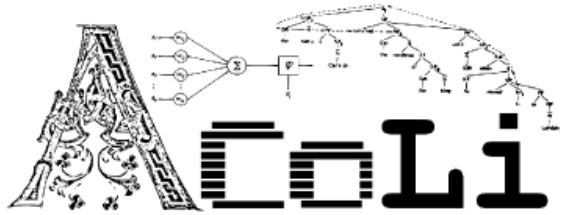
- correlations confirm GH core assumptions
- we can induce language-specific Givenness Hierarchies from coreference annotations
- impulses for the revision and practical application of the GH
  - **unique, referential** and **type** don't seem to differentiate well
    - **referential** and **type** don't actually differ in „givenness" in the sense of „hearer-old information", but rather in the speaker's intent
    - the identification of **unique** is heavily heuristic, this might just be ill-defined

Conclusion



- correlations confirm GH core assumptions
- we can induce language-specific Givenness Hierarchies from coreference annotations
- impulses for the revision and practical application of the GH
  - **unique, referential** and **type** don’t seem to differentiate well
  - because **unique** in the GH is a subset of **referential**, we suggest to abandon **unique** in favor of **(extended) referential** in future studies
    - the differentiation between **referential** and **type** is clear-cut (presence or lack of subsequent mention)

Conclusion



- correlations confirm GH core assumptions
- we can induce language-specific Givenness Hierarchies from coreference annotations
- impulses for the revision and practical application of the GH
- outlook
  - compare with manual annotations for GH
  - use coref-based GH training data to develop an automated (coreference-free) givenness tagger
  - explore theoretical and practical ramifications, also in comparison with alternative theories of information status

Thank you for your attention