The limits of accessibility

Geoffrey Haig [1]   Stefan Schnell [2]   Nils Schiborr [1]

[1]University of Bamberg

[2]Centre of Excellence for the Dynamics of Language & University of Melbourne

ALT12, A.N.U. Canberra
7 December 2017

Outline

Reference and accessiblity

Reference and accessibility

Ariel 1990[2014]; Fretheim & Gundel 1996 (eds.); Chafe 1976, 1994; among many others

- choice of form for referents (real or unreal entities talked about)
- speaker's assessment of addressee's accessibility to intended referent
- 1-to-1 relation in prominent models like Accessibility Theory

Discourse referents

(1) *A man and his wife are working in the garden. They are harvesting pumpkins.*

Discourse referents

(2) *Now they take the pumpkins with them. The woman is dragging their child along.*

Accessibility Theory: Principles of marking

less accessible:

- are more informative
- are more rigid / "uniquely referring"
- have greater phonological substance

The Accessibility Hierarchy

degree of accessibility		marker
lowest	**a.**	full name + modifier
	b.	full name
	c.	long definite description
	d.	short definite description
	e.	last name
	f.	first name
	g.	distal demonstrative + modifier
	h.	proximal demonstrative + modifier
	i.	distal demonstrative (+ NP)
	j.	proximal demonstrative (+ NP)
	k.	stressed pronoun + gesture
	l.	stressed pronoun
	m.	unstressed pronoun
	n.	cliticized pronoun
highest	**o.**	*extremely high accessibility markers* (gaps, *wh*-traces, reflexives, agreement)

Table 1 The accessibility marking scale Ariel (1990: 73 ex. 1).

→relative ranking of *language-specific* forms according to *universal* principle of accessibility

Our study

$\rightarrow$factors considered:

- ▶ newness: when a referent is introduced into the universe of discourse (first mention)

- ▶ distance between anaphor and antecedent

- ▶ humanness: human or anthropomorphised beings versus non-human

$\rightarrow$cross-corpus approach

Multi-CAST — Multilingual Corpus of Annotated Spoken Texts
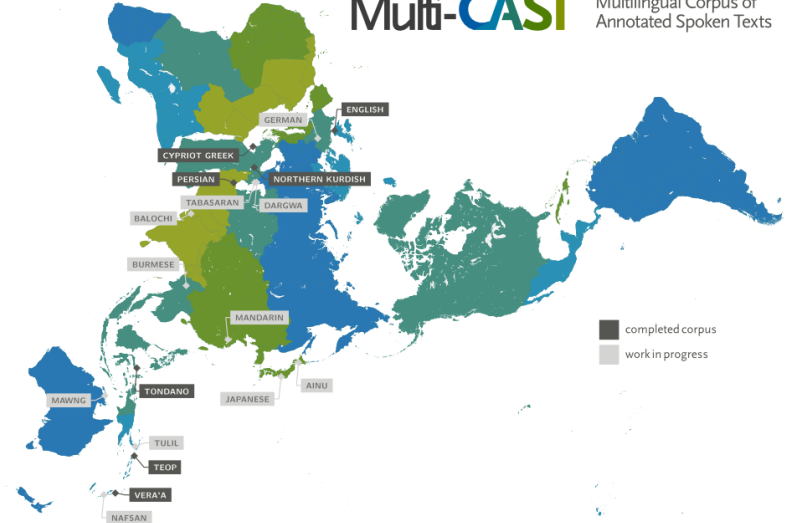
`https://lac.uni-koeln.de/de/multicast/`

Multi-CAST
Multiingual Corpus of Annotated Spoken Texts

- original texts, mainly narratives
- includes some Pear Film re-tellings (Persian)
- excluding conversational data at present

Multi-CAST: global overview

Multi-CAST: text corpora

language	corpus	text type	texts	length h:mm:ss	clause units
Cypriot Greek	*cypgreek*	traditional	3	—	1,071
English	*english*	autobiogr.	2	1:30:09	2,245
Northern Kurdish	*nkurd*	traditional	2	0:32:05	1,101
Persian	*persian*	stim.-based	29	0:52:32	1,417
Teop	*teop*	traditional	4	0:46:35	1,302
Tondano	*tondano*	auto./st.-b.	8	1:16:18	1,086
Vera'a	*veraa*	traditional	10	2:01:48	3,606
		collection totals	58	6:59:27	11,828

Corpus annotation: GRAID
Grammatical Relations and Animacy in Discourse

core annotations with GRAID:

- ▶ form: noun phrase, pronoun, zero
- ▶ person and animacy: 1st, 2nd, human, non-human
- ▶ syntactic function: S, A, P, other functions

→define these as maximally cross-linguistically applicable

GRAID example (simplified)

(3) Schiborr (2017: ex 17)

I	went	along	with	this old man,	Mr. Brown,	he	was
pro.1:s	v:pred		adp	np.h:obl		pro.h:s	cop

a nice old man.	__	Used to have	a team of four great
np.h:pred	0.h:a	v:pred	np:p

horses.	We	used to have to go	to	work	and	_	do
	pro.1:s	v:pred	adp	np:g		0.1:a	v:pred

the ploughing	with	'em,
np:p	adp	=wpro:obl

Corpus annotation: RefIND
Referent indices

RefIND: unique numerical identifier

- assigns each referent a unique index
- assigns each mention of that referent the same index
- exclude non-referential expressions
- accompanying referent list with standardised set of features (ontological category, . . .)

→no necessity to directly code derivative notions such as (discourse) topicality

RefIND example (simplified)

(4) Schiborr (2017: ex 17)

I	went	along	with	this old man,	Mr. Brown,	he	was
pro.1:s	v:pred		adp	np.h:obl		pro.h:s	cop
000				021		021	

a nice old man.	__	Used to have	a team of four great
np.h:pred	0.h:a	v:pred	np:p
	021		024

horses.	We	used to have to go	to	work	and	_	do
	pro.1:s	v:pred	adp	np:g		0.1:a	v:pred
	029					029	

the ploughing	with	'em,
np:p	adp	=wpro:obl
		024

Referent introduction

The challenge of managing discourse referents
Ariel 1990; Du Bois 1987, 2003a,b; Lambrecht 1994; Prince 1998

- ▶ according to activation-based mdels, referent introduction regarded as highly processing-costly
- ▶ assume that languages provide specialised syntactic positions to accommodate new referents
- ▶ Preferred Argument Structure (*predictable locus for unpredictable work*, Du Bois 2003a:78)
- ▶ also left-dislocation, nominal predicate, existential/presentational constructions, . . .

First mention

- first occurrence in a text (in its recording) (cf Ariel's *linguistic context*)

- differentiation between *(brand) new* versus *bridging*

- *brand new* should require more special treatment than *bridging*

$\rightarrow$we investigate patterns of referent introduction in four of our corpora: Cypriot Greek, English, Teop, Vera'a

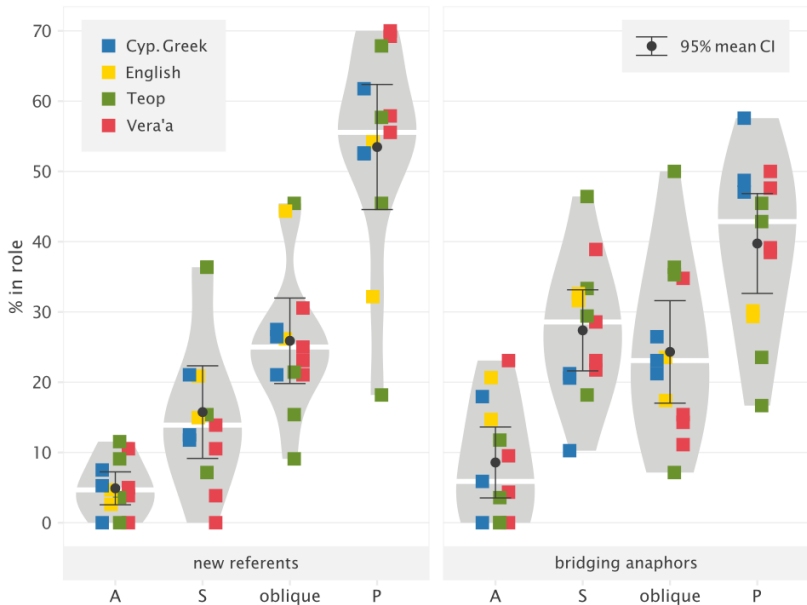# Referent introduction across four corpora

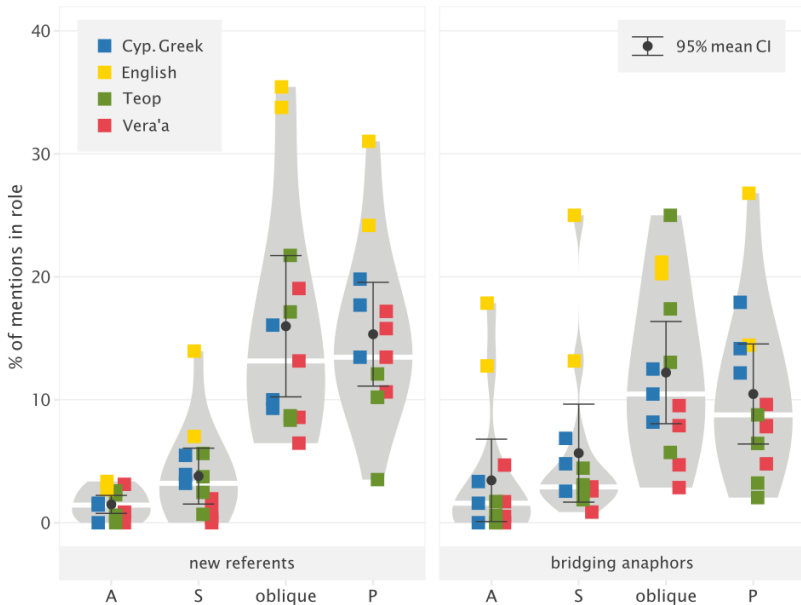Figure: 1: Where do new mentions go?

Referent introduction across four corpora

Figure: 2: Proportion of new mentions in each role

REFERENT TRACKING

Case study: Antecedent distance and Accessibility

Ariel 1990; Chafe 1976; Givón 2017; Schiborr 2017

- ▶ distance to antecedent co-determines Accessibility, reflected in choice of form

- ▶ Position on the Accessiblity Hierarchy claimed to vary monotonically according to distance: more distant →less accessible

- ▶ Additional investigation: role of Animacy [+/-human] in Accessibility

Data

- Spoken spontaneous narrative English (oral history)
- subset of the Multi-CAST English corpus (Schiborr 2016)
- Two narratives: 1265 clauses, 8100 words, approx. 1 hour
- Annotated with GRAID, and RefIND (Schiborr et al. 2017)

Procedure

- Annotated corpus imported into Multi-castR library (prototype Schiborr 2017)

- Observations: all referring expressions that have more than one mention in the corpus

- 696 data points, majority (63%) non-lexical: pronominal (58%), or zero (5%)

- Predictor variable: Distance to antecedent measured in clause units (winsorized at 20)

- Dependent variables: Form of anaphor (see next slide)

Form types considered

Simplified version of Accessibility Hierarchy (Ariel 1990)

		Expression type	Expression properties
	low	proper names	
		lexical noun phrases	heavy – short definite – demonstrative
		demonstratives	distal – proximal
←degree of accessibility →		personal pronouns	
		clitic pronouns	
		zero	
	high	[forced gaps, reflexives, etc.]	

Accessibility
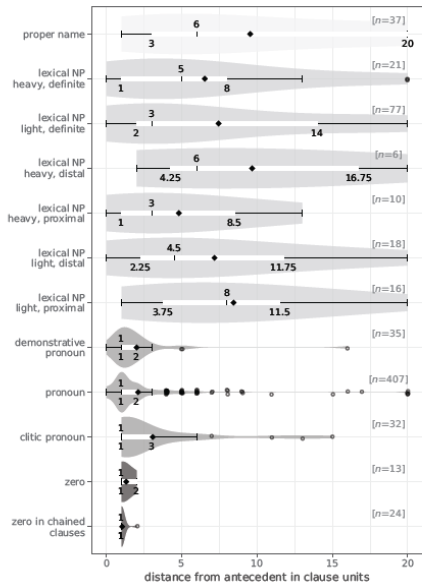
Exporing the interaction of distance and humanness in Accessibility

Prediction of Accessibility Theory:

- ▶ human referents are more salient, thus inherently higher in accessibility.

In our terms:

- ▶ If referent is +human (e.g. *that woman*) →
  high-accessibility forms (non-lexical, e.g. pronoun / zero) available as anaphora for longer distances

- ▶ If referent is -human (e.g. *my car*) →
  high-accessibility forms (non-lexical, e.g. pronoun / zero) available as anaphora for shorter distances

- ▶ Not previously tested systematically for spoken language data

Humanness
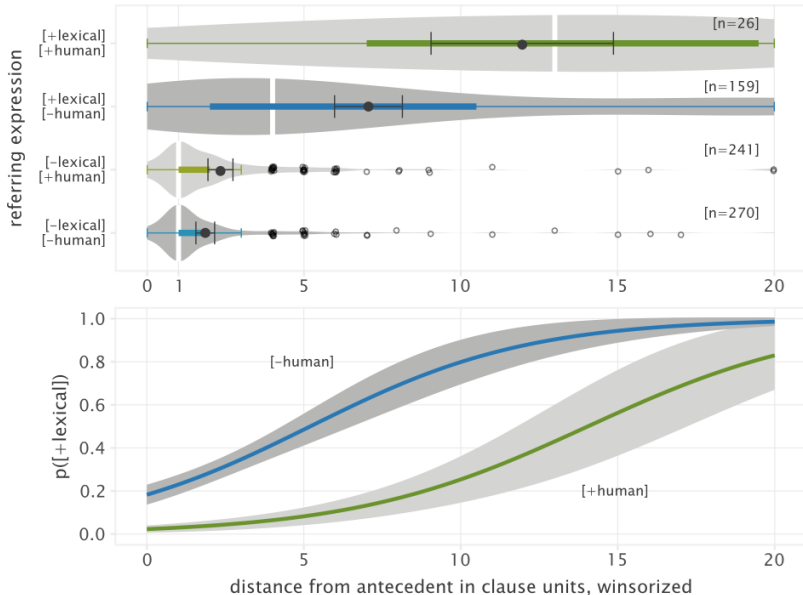
Conclusions and prospects

Conclusions

newness:

- ▶ referent introduction does not require or prefer a special locus, separate from other referents
- ▶ to the contrary, seems to prefer P role which is inherently linked to A (cf Primus 1997)
- ▶ Schnell (submitted): no interaction with *information pressure*, i.e. number of entities talked about in a text

tracking:

- ▶ "By and large, all non-lexical expressions pattern alike, as do all lexical expressions." (Schiborr 2017:64)
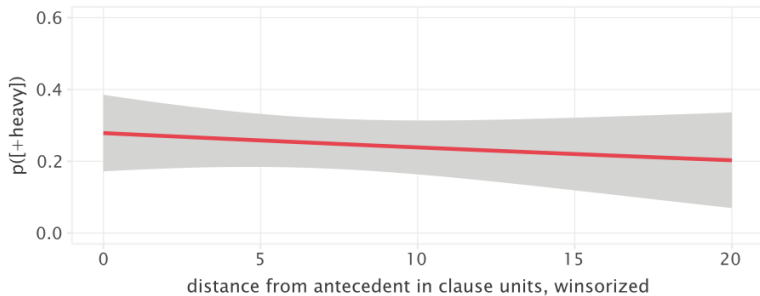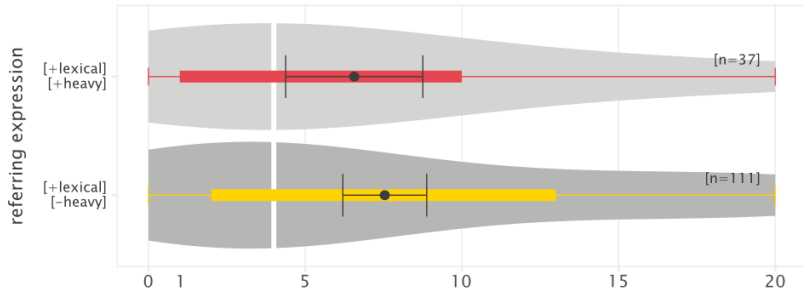
Prospect

future work:

- data-driven modelling of referential choice as alternative approach to hypothesis testing

→explanation would have to be psycho-linguistically testable

Weight

Accessibility