The limits of accessibility

Geoffrey Haig ¹  Stefan Schnell ²  Nils Schiborr ¹

¹University of Bamberg
²Centre of Excellence for the Dynamics of Language & University of Melbourne

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

Introduction: reference and accessibility

Multi-CAST: Multilingual Corpus of Annotated Spoken Texts

Referent introduction

Referent tracking

Conclusions and outlook
REFERENCE AND ACCESSIBILITY
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
(1) A man and his wife are working in the garden. They are harvesting pumpkins.
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

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

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

→ 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

→ cross-corpus approach
https://lac.uni-koeln.de/de/multicast/
Multi-CAST
Multi-lingual 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

<table>
<thead>
<tr>
<th>language</th>
<th>corpus</th>
<th>text type</th>
<th>texts</th>
<th>length h:mm:ss</th>
<th>clause units</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cypriot Greek</td>
<td>cypgreek</td>
<td>traditional</td>
<td>3</td>
<td>—</td>
<td>1,071</td>
</tr>
<tr>
<td>English</td>
<td>english</td>
<td>autobiogr.</td>
<td>2</td>
<td>1:30:09</td>
<td>2,245</td>
</tr>
<tr>
<td>Northern Kurdish</td>
<td>nkurd</td>
<td>traditional</td>
<td>2</td>
<td>0:32:05</td>
<td>1,101</td>
</tr>
<tr>
<td>Persian</td>
<td>persian</td>
<td>stim.-based</td>
<td>29</td>
<td>0:52:32</td>
<td>1,417</td>
</tr>
<tr>
<td>Teop</td>
<td>teop</td>
<td>traditional</td>
<td>4</td>
<td>0:46:35</td>
<td>1,302</td>
</tr>
<tr>
<td>Tondano</td>
<td>tondano</td>
<td>auto./st-b.</td>
<td>8</td>
<td>1:16:18</td>
<td>1,086</td>
</tr>
<tr>
<td>Vera’a</td>
<td>veraa</td>
<td>traditional</td>
<td>10</td>
<td>2:01:48</td>
<td>3,606</td>
</tr>
</tbody>
</table>

| collection totals   | 58      | 6:59:27       | 11,828|
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
I went along with this old man, Mr. Brown, he was a nice old man. Used to have a team of four great horses. We used to have to go to work and do the ploughing with ’em,
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
(4) Schiborr (2017: ex 17)

I went along with this old man, Mr. Brown, he was a nice old man. Used to have a team of four great horses. We used to have to go to work and do the ploughing with 'em,
Referent introduction
The challenge of managing discourse referents
Ariel 1990; Du Bois 1987, 2003a,b; Lambrecht 1994; Prince 1998

- according to activation-based models, referent introduction regarded as highly processing-costly
- assume that languages provide specialised syntactic positions to accommodate new referents
- also left-dislocation, nominal predicate, existential/presentational constructions, ...
First mention
Ariel 1990; Prince 1981

- first occurrence in a text (in its recording) (cf Ariel's linguistic context)
- differentiation between \((\text{brand}) \text{ new}\) versus \text{bridging}
- \text{brand new} should require more special treatment than \text{bridging}

→ 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 Accessibility 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)

<table>
<thead>
<tr>
<th>Expression type</th>
<th>Expression properties</th>
</tr>
</thead>
<tbody>
<tr>
<td>low</td>
<td></td>
</tr>
<tr>
<td>proper names</td>
<td></td>
</tr>
<tr>
<td>lexical noun phrases</td>
<td>heavy – short</td>
</tr>
<tr>
<td></td>
<td>definite – demonstrative</td>
</tr>
<tr>
<td>demonstratives</td>
<td>distal – proximal</td>
</tr>
<tr>
<td>personal pronouns</td>
<td></td>
</tr>
<tr>
<td>clitic pronouns</td>
<td></td>
</tr>
<tr>
<td>high</td>
<td></td>
</tr>
<tr>
<td>zero</td>
<td>[forced gaps, reflexives, etc.]</td>
</tr>
</tbody>
</table>
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*) $\rightarrow$
  high-accessibility forms (non-lexical, e.g. pronoun / zero) available as anaphora for longer distances

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

- Not previously tested systematically for spoken language data
Humanness

- [+lexical] [+human] (n=26)
- [+lexical] [-human] (n=159)
- [-lexical] [+human] (n=241)
- [-lexical] [-human] (n=270)

Distance from antecedent in clause units, winsorized
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)
future work:

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

→ explanation would have to be psycho-linguistically testable
Weight

The graph shows the distribution of weight for referring expressions with [+lexical] [+]heavy and [+lexical] [-heavy] over the distance from antecedent in clause units, winsorized. The distribution is represented by density plots with shaded areas indicating the range of values and a line with a point indicating the median. The sample sizes are [n=37] for [+lexical] [+]heavy and [n=111] for [+lexical] [-heavy].