Discourse coherence

- psycholinguistic and corpus-based research has focused on *discourse coherence*: tracking and manipulation of shared referents

- *production* — choice of different anaphoric devices
  
  (Kibrik 2011; Kibrik et al. 2016; Torres Cacoullos & Travis 2019; Schiborr, In prep.; etc.)

- *processing* — anaphora resolution
  
  (Kehler 2002; Sauermann & Gagarina 2017; Holler & Suckow 2016; etc.)

- comparatively less attention on the *introduction of new referents* into discourse
Accessibility and activation

- most research on new information is rooted in the **functional/typological tradition**
- usually involves some notion of **accessibility** or **activation**
  (Chafe 1976; Fox 1987; Lambrecht 1994; Givón 1983; Du Bois 1987; Ariel 1990, etc.)
new referents are understood to present a cognitive challenge for discourse participants

- production — introductions are heavier
- perception — increased competition between referents

hence:

Chafe’s (1976) “one new concept at a time” constraint
Du Bois’ (1987) quantity constraint
Syntactic adaptation

- key claim: syntax adapts to the challenge of introducing new referents

- Lambrecht (1994): principle of the separation of role and reference (PSRR)
  (see also Payne 1992)

- Du Bois (1987): subjects of intransitive verbs with vague semantics as entry-points for new referents

- Dixon (1987): specialized syntactic constructions (e.g. presentational) for new referents
This study

- here, we test these predictions systematically, based on the analysis of

A. uniformly annotated, spoken corpora from 10 languages
   (Multi-CAST, Haig & Schnell 2020; accessible online at multicast.aspra.uni-bamberg.de)

B. introductions in Pear Story retellings from 4 languages
   (cf. Chafe 1980; watch film at youtube.com/watch?v=bRNSTxTpG7U)
Preferred syntactic positions

PART 1a
◆ preferred morphosyntactic strategies for referent introduction:
evidence from Multi-CAST
Preferred syntactic positions

- little support for preferred status of intransitive subjects (S)

- instead: non-subject roles, especially objects (P), have highest proportions of new mentions
Preferred syntactic positions

PART 1b

- preferred morphosyntactic strategies for referent introduction:
  evidence from Pear Story retellings
(A) man picking pears

![Graph showing the distribution of introductions in position for different languages and elaboration levels. The graph includes categories such as subject trans. (A), object (P), subject intrans. (S), S-motion, existential, and other. Different languages are represented by different colors: English (blue), Mandarin (green), Persian (orange), and Vera’a (red). The graph also indicates whether the introduction is elaborated or not.]
(B) man leading goat

<table>
<thead>
<tr>
<th>Introduction in Position</th>
<th>English</th>
<th>Mandarin</th>
<th>Persian</th>
<th>Vera’a</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject trans. (A)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Object (P)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subject intrans. (S)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S-motion</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Existential</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Elaborated</th>
<th>No</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mandarin</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Persian</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vera’a</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
(C) boy stealing fruit
(D) girl riding bike

![Graph showing the percentage of introductions in different positions for English, Mandarin, Persian, and Vera’a, with elaborated and no/yes options.

- Subject trans. (A)
- Object (P)
- Subject intrans. (S)
- S-motion
- Existential
- Other

% of introductions

- English
- Mandarin
- Persian
- Vera’a

elaborated no yes

Introduction in position

0 20 40 60 80 100

SLE 2020 — Introducing new referents — Schnell, Haig, Schiborr, Vollmer — University of Bamberg et al.
(E) three boys

% of introductions

- subject trans. (A)
- object (P)
- subject intrans. (S)
- S-motion
- existential
- other

introduction in position

- English
- Mandarin
- Persian
- Vera’a

elaborated

- yes
- no

SLE 2020 — Introducing new referents — Schnell, Haig, Schiborr, Vollmer — University of Bamberg et al.
Preferred syntactic positions

- **introductions in S fairly frequent** in Pear Stories, especially with **motion predicates**
- **often semantically enriched**, e.g. by further adjuncts *(a boy *on a bike*, *a man leading a goat*)
Presentational constructions

- dedicated presentational constructions confined to specific local contexts
  e.g. at the beginning of a text or scene

- elsewhere not common, with the exception of Mandarin

- possible other motivations: “out of the blue” appearances
  (cf. Chafe 1980; Du Bois 2004a, b)
PART 2

- separation of new information from general content advancement:
  evidence from Multi-CAST
Separation of new information

- little evidence for the separation of introductions from the flow of narrative content
  i.e. most introductions not focused on through structural or semantic isolation

- instead:
  most introductions accompanied by given referents in the same clause
Linking new information

- How come P sticks out as a locus for new information?
- Hypothesis: *transitive constructions provide a structural scaffold for linking new referents* (in P) *to already established ones* (in A)
- Other types of linking: predicates of perception, possessive constructions
PART 3

- (intransitive) predicates used for introductions tend to be semantically void:
  evidence from Multi-CAST
![Graph showing the percentage of introductions in Mandarin and N. Kurdish for different predicate types. The x-axis represents predicate types: effect, activity, motion, state, speech, experience, perception, possession, and other. The y-axis represents the percentage of introductions. The graph shows the comparison between Mandarin (green) and N. Kurdish (orange) for linking to given information (no, yes).]
Semantically void predicates

- Intransitive predicates involved in introductions are not obviously semantically void.

- State predicates only account for between a fifth and a third of all introductions: more than motion predicates, but much fewer than change-of-state transitives.

- By far the most frequent predicate types are transitive.

- Intransitives often motion events, and often semantically enriched.
No specialization

- high frequency of introductions in object and other non-subject positions
- otherwise no obvious generalization re: special status of any role or predicate type
- instead: variation between text types (Multi-CAST vs. Pear Stories) and between languages
Integration

- interlocutors focus on conceptual content and states of affairs

- new information is seamlessly integrated into the narrative flow (with possible exception of major episodic breaks)
Role profiles

- **introductions in A are avoided**
due to the convergence of humanness and topicality
  (Haig & Schnell 2016; Schnell et al. to appear; Haig et al. 2020)

- **S is semantically broad**
  and overall very frequent, but not specifically associated with introductions

- **P and non-subject roles**
  are naturally associated with semantic roles that link to new information (e.g. perception, creation)
Cognitive demands of introductions

- whether introductions are cognitively more demanding cannot be conclusively surmised from our results
- referent introduction may be potentially challenging, but this need not be reflected in linguistic structure
- introductions are much rarer than given mentions (1 in 10)
- possible exception are transitive constructions, but here solution is not to separate, but to link
Cognitive demands of introductions

- are introductions actually more demanding than keeping track of old information?
- for the speaker, new information is not actually new
- inverse perspective: given information easier to deal with than new
- notion of challenging introductions thus consequence of focus on scene-setting contexts
Summary: Referent introduction

- referent introduction is less disruptive to syntax than has been claimed
  (Lambrecht 1994; Payne 1992; Du Bois 1987; Dixon 1987)

- we have found little evidence for
  - the separation of new information from content advancement;
  - the isolation of new information in intransitive constructions; and
  - the association of new information with vague verb semantics

- use of presentational constructions common at episode breaks, but not limited to new referents (Abbott 1992, 1993)
Summary: Referent introduction

- instead: referent introduction is primarily content-driven
- speakers seamlessly integrate new information into existing syntactic schemas, without recourse to specialized constructions
- our findings also cast doubt on the claim that referent introduction is cognitively demanding (or at least more than other aspects of discourse)
APPENDICES
Multi-CAST: The corpora

- **non-elicited, monologic spoken narratives**
from the **Multi-CAST collection**  (Haig & Schnell 2020)

<table>
<thead>
<tr>
<th>Language</th>
<th>Language Family</th>
<th>Research Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cypriot Greek</td>
<td>Indo-European, Greek</td>
<td>(Hadjidas &amp; Vollmer 2015)</td>
</tr>
<tr>
<td>English</td>
<td>Indo-European, Germanic</td>
<td>(Schiborr 2015)</td>
</tr>
<tr>
<td>Mandarin</td>
<td>Sino-Tibetan, Sinitic</td>
<td>(Vollmer 2020)</td>
</tr>
<tr>
<td>Nafsan</td>
<td>Austronesian, Oceanic</td>
<td>(Thieberger &amp; Brickell 2019)</td>
</tr>
<tr>
<td>Northern Kurdish</td>
<td>Indo-European, Iranian</td>
<td>(Haig et al. 2019)</td>
</tr>
<tr>
<td>Sanzhi Dargwa</td>
<td>Nakh-Daghest., Dargwa</td>
<td>(Forker &amp; Schiborr 2019)</td>
</tr>
<tr>
<td>Tabasaran</td>
<td>Nakh-Daghest., Lezgic</td>
<td>(Bogomolova &amp; Schiborr, In prep.)</td>
</tr>
<tr>
<td>Teop</td>
<td>Austronesian, Oceanic</td>
<td>(Mosel &amp; Schnell 2015)</td>
</tr>
<tr>
<td>Tulil</td>
<td>Papuan, Taulil-Butam</td>
<td>(Meng 2019)</td>
</tr>
<tr>
<td>Vera’a</td>
<td>Austronesian, Oceanic</td>
<td>(Schnell 2015)</td>
</tr>
</tbody>
</table>
Multi-CAST: Annotations

- the corpora have been annotated for
  - the form and role of referring expressions, (with GRAID, Haig & Schnell 2014)
  - co-reference relations between mentions (with RefIND, Schiborr et al. 2018), and
  - semantic predicate types (with PredSem, Haig et al. In prep.)
## Multi-CAST: The sample

<table>
<thead>
<tr>
<th>corpus</th>
<th>clause units</th>
<th>sampled referents</th>
<th>sampled mentions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cypriot Greek</td>
<td>1 071</td>
<td>165</td>
<td>1 665</td>
</tr>
<tr>
<td>English</td>
<td>4 184</td>
<td>918</td>
<td>5 380</td>
</tr>
<tr>
<td>Mandarin</td>
<td>1 194</td>
<td>198</td>
<td>1 641</td>
</tr>
<tr>
<td>Nafsan</td>
<td>1 012</td>
<td>172</td>
<td>1 393</td>
</tr>
<tr>
<td>Northern Kurdish</td>
<td>1 359</td>
<td>181</td>
<td>2 420</td>
</tr>
<tr>
<td>Sanzhi Dargwa</td>
<td>1 066</td>
<td>178</td>
<td>1 284</td>
</tr>
<tr>
<td>Tabasaran</td>
<td>1 386</td>
<td>190</td>
<td>2 016</td>
</tr>
<tr>
<td>Teop</td>
<td>1 302</td>
<td>143</td>
<td>1 620</td>
</tr>
<tr>
<td>Tulil</td>
<td>1 264</td>
<td>230</td>
<td>2 111</td>
</tr>
<tr>
<td>Vera’a</td>
<td>3 608</td>
<td>428</td>
<td>5 347</td>
</tr>
<tr>
<td><strong>totals</strong></td>
<td><strong>17 446</strong></td>
<td><strong>2 803</strong></td>
<td><strong>24 877</strong></td>
</tr>
</tbody>
</table>
Multilingual Corpus of Annotated Spoken Texts

multicast.aspra.uni-bamberg.de/
**Pear Stories**

- **narrative retellings of the Pear Film** *(see Chafe 1980)*

<table>
<thead>
<tr>
<th>Language</th>
<th>Family</th>
<th>Number of Texts</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>Indo-European, Greek</td>
<td>20 texts</td>
<td>(from Chafe 1980)</td>
</tr>
<tr>
<td>Mandarin</td>
<td>Sino-Tibetan, Sinitic</td>
<td>20 texts</td>
<td>(from Erbaugh 2001)</td>
</tr>
<tr>
<td>Persian</td>
<td>Indo-European, Iranian</td>
<td>29 texts</td>
<td>(from Adibifar 2016)</td>
</tr>
<tr>
<td>Vera’a</td>
<td>Austronesian, Oceanic</td>
<td>12 texts</td>
<td></td>
</tr>
</tbody>
</table>


References


References


Hadjidas, Harris & Vollmer, Maria C. 2015. Multi-CAST Cypriot Greek. In Haig, Geoffrey & Schnell, Stefan (eds.), Multi-CAST.


References


References


**References**


