SLE 2020

Introducing new referents

A corpus-based cross-linguistic perspective

Stefan Schnell
University of Bamberg / ARC CoEDL

Geoffrey Haig
University of Bamberg

Nils Norman Schiborr
University of Bamberg

Maria Carina Vollmer
University of Freiburg / Australian National University

26 August 2020
v1.0

DFG

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.)

Cognitively challenging

◆ 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):
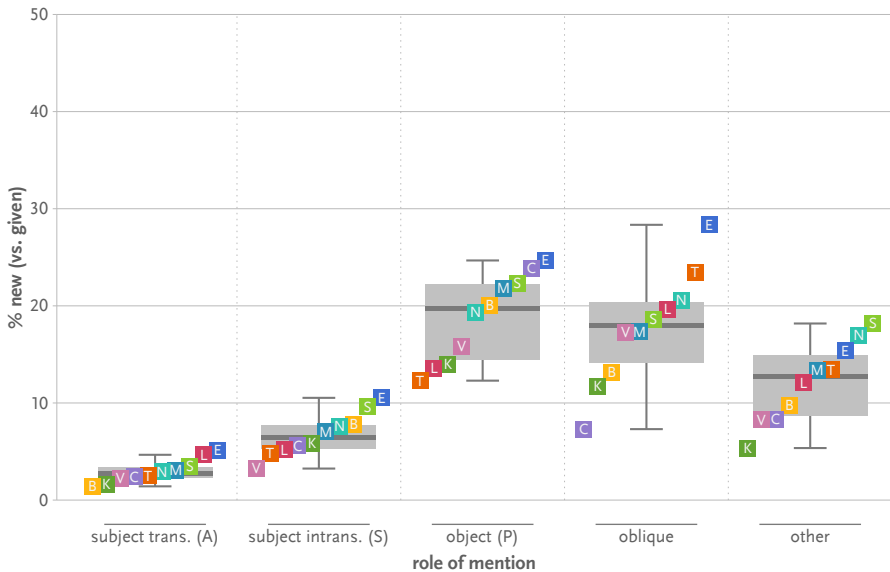 specialized syntactic constructions (e.g. presentationals)
  for new referents

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

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

7

% new (vs. given) vs. **role of mention**

Categories: subject trans. (A), subject intrans. (S), object (P), oblique, other

Legend:
- C = C. Greek
- E = English
- M = Mandarin
- N = Nafsan
- K = N. Kurdish
- S = S. Dargwa
- B = Tabasaran
- T = Teop
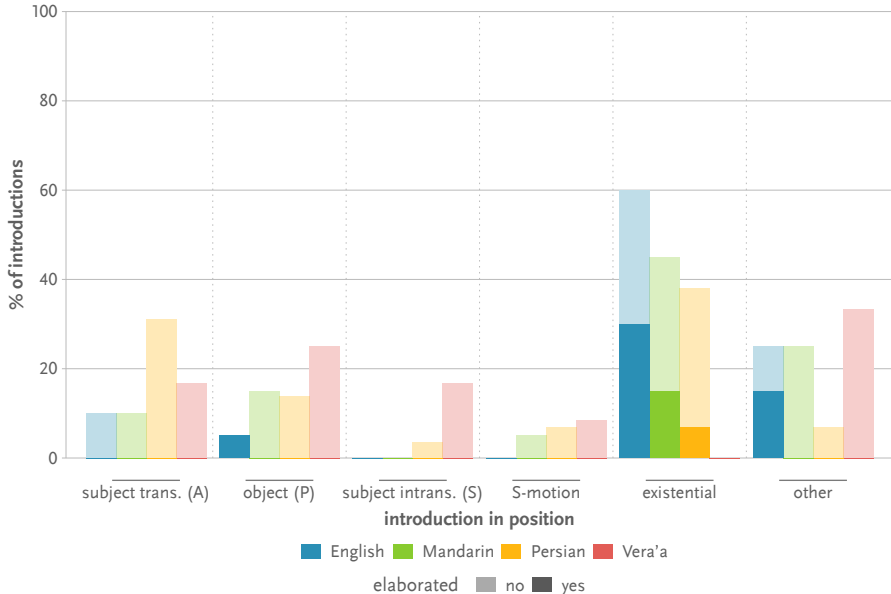- L = Tulil
- V = Vera'a

Preferred syntactic positions

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

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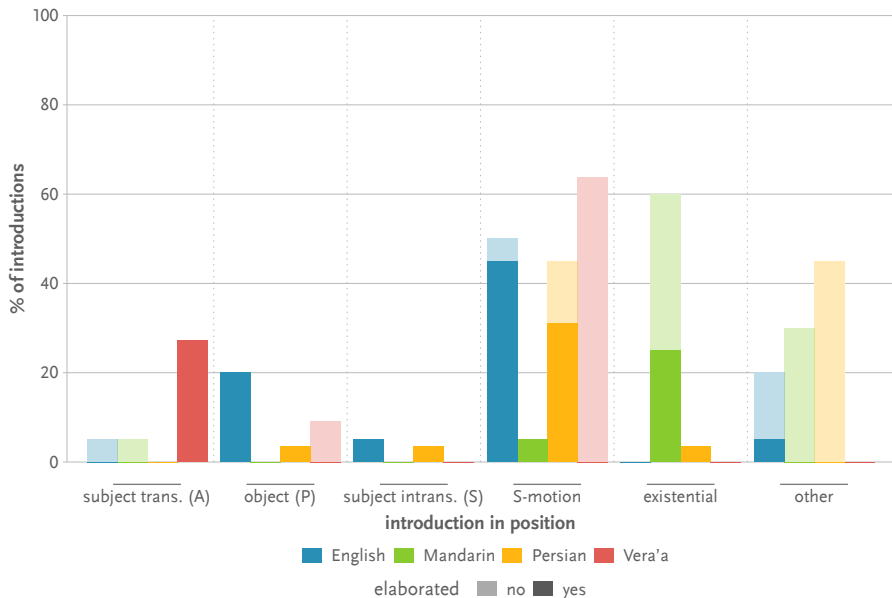 **have highest proportions of new mentions**

Preferred syntactic positions

PART 1b

◆ **preferred morphosyntactic strategies for referent introduction:**
evidence from Pear Story retellings

(A) man picking pears

% of introductions (y-axis): 0, 20, 40, 60, 80, 100

introduction in position (x-axis): subject trans. (A), object (P), subject intrans. (S), S-motion, existential, other

Legend: English, Mandarin, Persian, Vera'a

elaborated: no, yes

(B) man leading goat

(C) boy stealing fruit

(D) girl riding bike

% of introductions vs. introduction in position

Legend: English, Mandarin, Persian, Vera'a

elaborated: no, yes
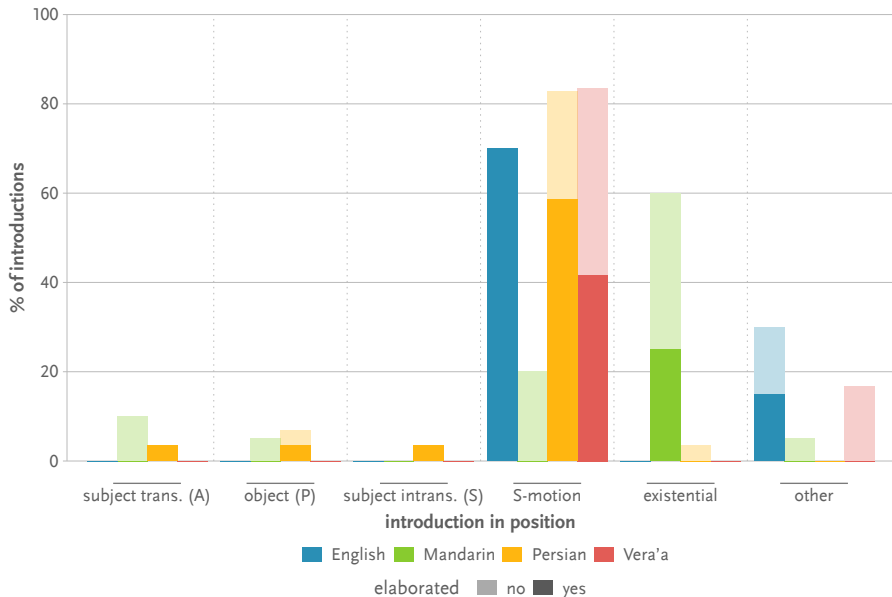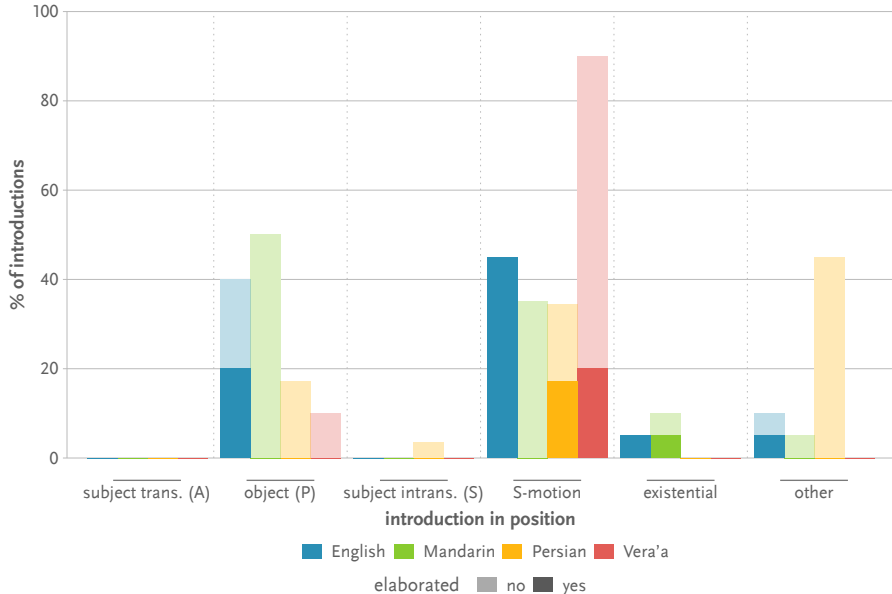
(E) three boys
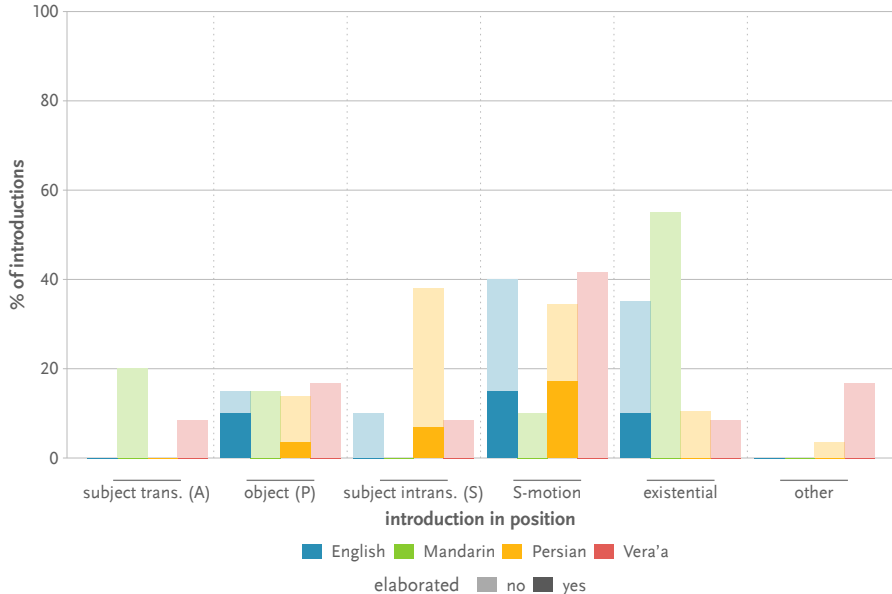
Preferred syntactic positions

- **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

% of clauses with introductions (y-axis, 0–100)

number of given mentions in clause (x-axis: 0, 1, 2, 3+)
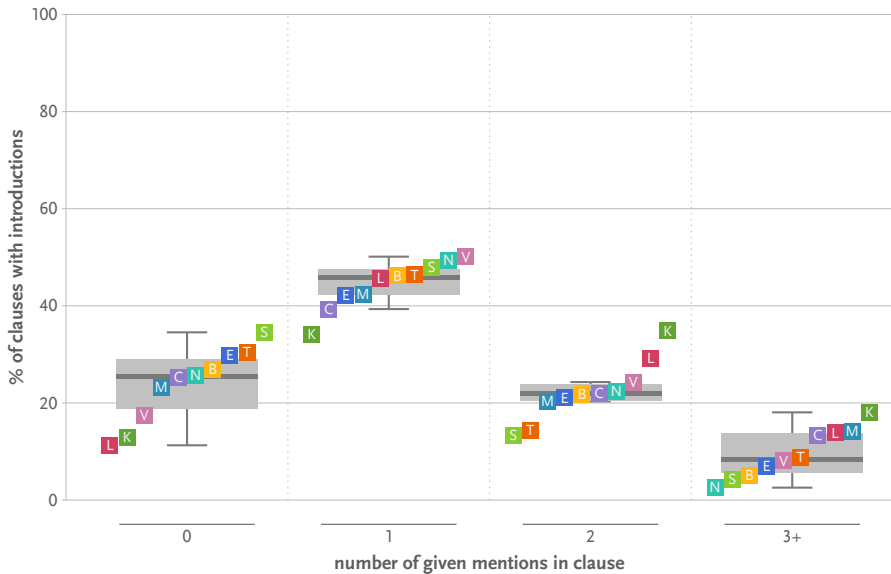
Legend:
- **C** C. Greek
- **M** Mandarin
- **K** N. Kurdish
- **B** Tabasaran
- **L** Tulil
- **E** English
- **N** Nafsan
- **S** S. Dargwa
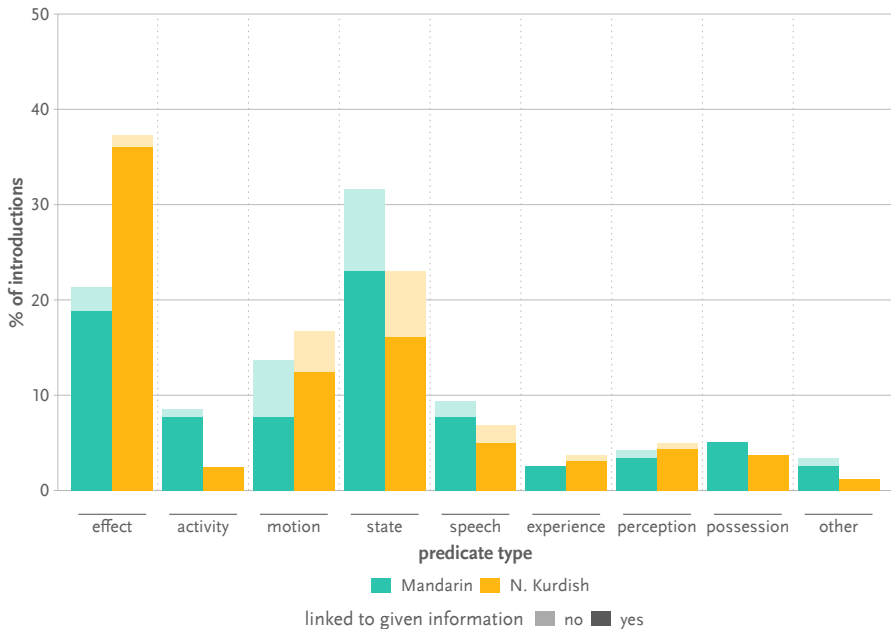- **T** Teop
- **V** Vera'a

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

Semantically void predicates

PART 3

- ◆ **(intransitive) predicates used for introductions tend to be semantically void:**
  evidence from Multi-CAST

% of introductions vs *predicate type*

Legend: Mandarin (teal), N. Kurdish (orange)

linked to given information — no (light), yes (dark)

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

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

32

Multi-CAST: The corpora

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

Cypriot Greek	Indo-European, Greek	(Hadjidas & Vollmer 2015)
English	Indo-European, Germanic	(Schiborr 2015)
Mandarin	Sino-Tibetan, Sinitic	(Vollmer 2020)
Nafsan	Austronesian, Oceanic	(Thieberger & Brickell 2019)
Northern Kurdish	Indo-European, Iranian	(Haig et al. 2019)
Sanzhi Dargwa	Nakh-Daghest., Dargwa	(Forker & Schiborr 2019)
Tabasaran	Nakh-Daghest., Lezgic	(Bogomolova & Schiborr, In prep.)
Teop	Austronesian, Oceanic	(Mosel & Schnell 2015)
Tulil	Papuan, Taulil-Butam	(Meng 2019)
Vera'a	Austronesian, Oceanic	(Schnell 2015)

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

corpus	clause units	sampled referents	sampled mentions
Cypriot Greek	1 071	165	1 665
English	4 184	918	5 380
Mandarin	1 194	198	1 641
Nafsan	1 012	172	1 393
Northern Kurdish	1 359	181	2 420
Sanzhi Dargwa	1 066	178	1 284
Tabasaran	1 386	190	2 016
Teop	1 302	143	1 620
Tulil	1 264	230	2 111
Vera'a	3 608	428	5 347
totals	**17 446**	**2 803**	**24 877**

Multi-CAST

Multilingual Corpus of Annotated Spoken Texts

`multicast.aspra.uni-bamberg.de/`

Pear Stories

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

English	Indo-European, Greek	20 texts	(from Chafe 1980)
Mandarin	Sino-Tibetan, Sinitic	20 texts	(from Erbaugh 2001)
Persian	Indo-European, Iranian	29 texts	(from Adibifar 2016)
Vera'a	Austronesian, Oceanic	12 texts	

References

Abbott, Barbara. **1992**. Definiteness, existentials, and the 'list' interpretation. In *SALT II* 19(1). 39–55.

Abbott, Barbara. **1993**. A pragmatic account of the definiteness effect in existential sentences. In Barker, Chris & Dowty, David R. (eds.), *Journal of Pragmatics: Proceedings of the Second Conference on Semantics and Linguistic Theory*, 1–16. Columbus, OH: Ohio State University.

Adibifar, Shirin. **2016**. Multi-CAST Persian. In Haig, Geoffrey & Schnell, Stefan (eds.), *Multi-CAST*.

Ariel, Mira. **1990**. *Accessing noun-phrase antecedents*. London: Routledge.

Arnold, Jennifer E. **2003**. Multiple constraints on reference form: Null, pronominal, and full reference in Mapudungun. In Du Bois, John & Kumpf, Lorraine & Ashby, William J. (eds.), *Preferred argument structure: Grammar as architecture for function,* 225–245. Amsterdam: John Benjamins.

Bogomolova, Natalia & **Schiborr**, Nils N. **In preparation**. Multi-CAST Tabasaran. In Haig, Geoffrey & Schnell, Stefan (eds.), *Multi-CAST*.

Chafe, Wallace. **1976**. Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In Li, Charles N. (ed.), *Subject and topic*, 25–55. New York: Academic Press.

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

38

References

Chafe, Wallace. **1980**. *The Pear Stories: Cognitive, cultural, and linguistic aspects of narrative production*. Norwood, NJ: Ablex.

Dixon, R. M. W. **1987**. Studies in ergativity. *Lingua* 71(1). 1–16.

Du Bois, John. **1987**. The discourse basis of ergativity. *Language* 63(4). 805–855.

Du Bois, John. **2003a**. Argument structure: Grammar in use. In Du Bois, John & Kumpf, Lorraine & Ashby, William J. (eds.), *Preferred argument structure: Grammar as architecture for function*, 11–60. Amsterdam: John Benjamins.

Du Bois, John. **2003b**. Discourse and grammar. In Tomasello, Michael (ed.), *The new psychology of language: Cognitive and function approaches to language structure*, Volume 2, 47–88. Mahwah, NJ: Erlbaum.

Erbaugh, Mary S. **2001**. *The Chinese Pear Stories: Narratives across seven Chinese dialects*. (http://pearstories.org/docu/ThePearStories.htm)

Forker, Diana & **Schiborr**, Nils N. **2019**. Multi-CAST Sanzhi Dargwa. In Haig, Geoffrey & Schnell, Stefan (eds.), *Multi-CAST*.

Fox, Barbara. **1987**. *Discourse structure and anaphora*. Cambridge: Cambridge University Press.

Givón, Talmy (ed.). **1983**. *Topic continuity in discourse*. Amsterdam: John Benjamins.

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

39

References

Gundel, Jeanette K. & **Hedberg**, Nancy & **Zacharski**, Ron. **1993**. Cognitive status and the form of referring expressions in discourse. *Language* 69(2). 274–307.

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

Haig, Geoffrey & **Schnell**, Stefan. **2014**. *Annotations using GRAID (Grammatical Relations and Animacy in Discourse): Introduction and guidelines for annotators.* Version 7.0. (https://multicast.aspra.uni-bamberg.de/#annotations)

Haig, Geoffrey & **Schnell**, Stefan. **2020**[2015]. *Multi-CAST: Multilingual Corpus of Annotated Spoken Texts.* (multicast.aspra.uni-bamberg.de)

Haig, Geoffrey & **Schnell**, Stefan. **2016**. *The discourse basis of ergativity revisited.* *Language* 92(3). 591–618.

Haig, Geoffrey & **Schnell**, Stefan & **Schiborr**, Nils N. **In preparation**. *PredSem: Annotations for semantic predicate types in natural discourse.*

Haig, Geoffrey & **Schiborr**, Nils N. & **Schnell**, Stefan. **2020**. On potential statistical universals of grammar in discourse: Evidence from Multi-CAST. Paper presented at the workshop Corpus-based typology: Spoken language from a cross-linguistic, perspective as part of the 42nd Annual Conference of the German Linguistic Society (DGfS 2020), Hamburg, Germany, 4–6 March 2020.

References

Haig, Geoffrey & **Vollmer**, Maria & **Thiele**, Hanna. **2019**. Multi-CAST Northern Kurdish. In Haig, Geoffrey & Schnell, Stefan (eds.), *Multi-CAST*.

Holler, Anke & **Suckow**, Katja. **2016**. *Empirical perspectives on anaphora resolution*. Berlin: Mouton de Gruyter.

Kehler, Andrew. **2002**. *Coherence, reference, and the theory of grammar*. Stanford: CSLI Publications.

Kibrik, Andrej A. **2011**. *Reference in discourse*. Oxford: Oxford University Press.

Kibrik, Andrej A. & **Khudyakova**, Mariya V. & **Dobrov**, Grigory B. & **Linnik**, Anastasia & **Zalmanov**, Dmitrij A.**2016**. Referential choice: Predictability and its limits. *Frontiers in Psychology* 7(1429).

Kumagai, Yoshiharu. **2006**. Information management in intransitive subjects: Some implications for the preferred argument structure theory. *Journal of Pragmatics* 38(5). 670–694.

Lambrecht, Knud. **1994**. *Information structure and sentence form: Topic, focus, and the mental representation of discourse referents*. Cambridge: Cambridge University Press.

Meng, Chenxi. **2019**. Multi-CAST Tulil. In Haig, Geoffrey & Schnell, Stefan (eds.), *Multi-CAST*.

References

Mosel, Ulrike & **Schnell**, Stefan. **2015**. Multi-CAST Teop. In Haig, Geoffrey & Schnell, Stefan (eds.), *Multi-CAST*.

Prince, Ellen F. **1981**. Toward a taxonomy of given-new information. In Cole, Peter (ed.), *Radical pragmatics*, 223–255. New York: Academic Press.

Schiborr, Nils N. **In Preparation**. *Lexical anaphora: A corpus-based typological study of referential choice*. PhD dissertation, University of Bamberg.

Schiborr, Nils N. **2015**. Multi-CAST English. In Haig, Geoffrey & Schnell, Stefan (eds.), *Multi-CAST*.

Schiborr, Nils N. & **Schnell**, Stefan & **Thiele**, Hanna. **2018**. *RefIND — Referent Indexing in Natural-language Discourse: Annotation guidelines.* Version 1.1. University of Bamberg. (https://multicast.aspra.uni-bamberg.de/#annotations)

Schnell, Stefan. **2015**. Multi-CAST Vera'a. In Haig, Geoffrey & Schnell, Stefan (eds.), *Multi-CAST*.

Schnell, Stefan & **Schiborr**, Nils N. & **Haig**, Geoffrey. **To appear**. Efficiency in discourse processing: Does morphosyntax adapt to accommodate new referents? To appear in *Linguistics Vanguard*, special issue on efficiency in language edited by Natalia Levshina and Steve Moran.

References

Sauermann, Antje & **Gagarina**, Natalia. **2017**. Parallelism influences ambiguous pronoun resolution in German. *Frontiers in Psychology* 8(1205).

Thieberger, Nick & **Brickell**, Timothy. **2019**. Multi-CAST Nafsan. In Haig, Geoffrey & Schnell, Stefan (eds.), *Multi-CAST*.

Torres Cacoullos, Rena & **Travis**, Catherine. **2019**. Variationist typology: Shared probabilistic constraints across (non-)null subject languages. *Linguistics* 57(3). 653–692.

Vollmer, Maria C. **2020**. Multi-CAST Mandarin. In Haig, Geoffrey & Schnell, Stefan (eds.), *Multi-CAST*.