

SLE 2020

Introducing new referents: A corpus-based cross-linguistic perspective

Stefan Schnell

University of Bamberg / ARC Centre of Excellence for the Dynamics of Language

Geoffrey Haig

University of Bamberg

Nils Norman Schiborr

University of Bamberg

Maria Vollmer

University of Freiburg / Australian National University

26 August 2020

01 — Title

Welcome to this pre-recorded talk on the introduction of new referents into discourse. My name is [insert name]. An online discussion round is scheduled for Thursday, August 27, from 9:30 to 9:40 (GMT+2). Please see the conference website for details and updates.

02 — Discourse coherence

A large body of research in psycholinguistics and corpus-based linguistics has focused on the notion of discourse coherence, that is the tracking and manipulation of referential entities assumed to be shared between interlocutors, or "given" in the sense of Gundel, Hedberg, and Zacharsky (1993). On the production side of things, studies have examined speakers' choices of different anaphoric devices, such as zero anaphors vs. pronouns, while on the processing side, the focus has been on establishing co-reference relations.

However, compared to the attention paid to given information, there has so far been relatively little research, in psycholinguistics and in corpus linguistics both, on the introduction of new information into discourse.

03 — Accessibility and Activation

Existing research on introductions is mainly rooted in the functional-typological tradition, and usually espouses some version of an accessibility or activation-based theory.

04 — Cognitively challenging

According to this view, introductions of new referents represent a cognitively challenging task for discourse participants. In terms of production, introductions tend to be associated with heavier, more informative forms, that is full NPs rather than pronouns or zero, and in terms of perception, introductions add to the overall cognitive load and exacerbate competition between referents in short-term memory. Hence, we get Chafe's "one new concept at a time" constraint and Du Bois' quantity constraint.

05 — Syntactic adaptation

The key prediction of accessibility and activation-based approaches is that syntax adaptively responds to the challenge of introducing new referents into discourse. The operative principle involves the separation of information management from content advancement in syntax, as per Lambrecht's principle of the separation of role and reference.

Two specific predictions are regularly made: One, that there is a favoured association of the subjects of intransitive verbs, especially those with vague semantics, with the role of serving as entry-points for new referents; and two, that certain specialized syntactic constructions, most notably presentationals/existentials, are specifically leveraged for introductions.

06 — This study

In the study presented here, we test these predictions systematically based on the statistical analysis of two data sets: The first consists of uniformly annotated, spoken corpora from ten genetically diverse languages, taken from the Multi-CAST collection; the second is a pilot study of new introductions in sets of Pear Story retellings from four languages.

Methodological overviews of the data used for this study are appended to the end of this presentation. Additional information and related publications for Multi-CAST can be found on the website. If you are unfamiliar with the Pear Film, feel free to pause this recording and watch it at the link provided here.

07 — Preferred syntactic positions

We begin by looking at evidence from the Multi-CAST corpora. The texts in Multi-CAST are mostly spontaneously produced traditional narratives – folktales, oral history, etc.

These data include only referents that are mentioned at least twice in a given text. We consider the first mention of a particular referent in a text to be its introduction; any subsequent mentions of that referent are given.

08 — Newness by role

[GRAPHIC: % new (vs. given) by role of mention in the Multi-CAST data]

Shown here are the proportions of new referents among all mentions in various syntactic positions: subjects of transitive clauses (A), subjects of intransitive clauses (S), objects (P), oblique arguments such as goals, instrumentals, and so on, and, lastly, all other positions including adjuncts and NP-internal modifiers such as possessors, etc.

09 — Preferred syntactic positions

As the data from Multi-CAST suggest, there is little evidence for a preferred position of intransitive subjects (S) for new information: Only a small fraction of the mentions in this position are new. Instead, we find that cross-linguistically, introductions make up a substantial proportion of mentions in various non-subject roles, among which direct objects (P) appear to be the most likely to host new information.

10 — Preferred syntactic positions

Next, we look at the corresponding picture in the Pear Story retellings. Unlike the Multi-CAST narratives, the Pear Story retellings are inherently controlled for content. As such, we focus our investigation on the five most narratively pertinent human referents in the film: the man picking pears, the man leading a goat, the boy with a bike who steals a basket of pears, the girl riding a bike, and the three boys. We will briefly look the way each of these five referents is introduced into discourse.

11 — Pear Story retellings

[GRAPHIC: % of introductions by position, man picking pears in the Pear Stories]

First off, the man picking pears, with whom the film starts. Presentational/existential constructions are quite common here, except in Vera'a, as are introductions in transitive clauses.

12 — Pear Story retellings

[GRAPHIC: % of introductions by position, man leading goat in the Pear Stories]

Next, the man leading a goat...

13 — Pear Story retellings

[GRAPHIC: % of introductions by position, boy stealing fruit in the Pear Stories]

...and the main character of the film, the boy stealing a basket of pears. Both of these referents arrive on the scene coming towards the camera, which is reflected in the strong preference for introductions as the subject of a motion predicate, often with additional elaboration, *a boy on a bike*. This pattern shows in three of the four languages, the exception being Mandarin.

14 — Pear Story retellings

[GRAPHIC: % of introductions by position, girl riding bike in the Pear Stories]

After stealing the basket of pears, the boy encounters a girl on the road, riding her bike. Here as well, introductions with motion predicates are very common, but so are introductions in object position, often with the verbs *meet*, *ride by*, or *see*. Existentials and introductions in the A role are almost completely absent.

15 — Pear Story retellings

[GRAPHIC: % of introductions by position, three boys in the Pear Stories]

And lastly, the three boys, which show a noticeably mixed picture.

16 — Preferred syntactic positions

In sum, introductions in S are fairly frequent in Pear Story retellings, especially with motion predicates. These introductions are often semantically enriched, for instance by elaboration with further adjuncts: *a boy on a bike*, *a man leading a goat*, etc.

17 — Presentational constructions

As for dedicated presentational constructions, these appear to be largely confined to specific local contexts, such as the beginning of texts, scenes, of episodes; as seen here, they are most common with the pear picker with whom the film opens. Elsewhere, presentationals are not common, with the exception of Mandarin, where they are frequent in contexts where the other three corpora employ motion predicates.

Another possible motivation for the use of presentationals could be "out of the blue" appearances, such as the rock in the Pear Film, which is only shown the very moment it becomes narratively pertinent.

18 — Separation of new information

Next, let us examine whether new referents are indeed separated from the general advancement of the narrative content. For this, we will once again leverage the data from Multi-CAST.

19 — New alongside given information

[GRAPH: % of clauses with introductions by number of given mentions in the clause in Multi-CAST]

These graphs show the number of mentions of given referents, that is those that were previously introduced, in the same clause as new referents in the ten Multi-CAST corpora.

As seen here, less than a third of new referents occur in isolation from given information in their clause; the vast majority are instead introduced alongside at least one already established referent, many with multiple.

20 — Separation of new information

These figures show that there is little evidence for the purported separation of new information from the general advancement of the narrative. The majority of introductions are not focused on through structural or semantic isolation, but are instead accompanied by one or more given referents in the same clause, anchoring them to the flow of the narrative.

21 — Linking new information

Looking back to beginning of the talk, how come objects stick out as a common locus for new information? Our hypothesis is that transitive constructions provide a structural scaffold for linking up new referents in the P role to already established referents in the A role. This allows new information to be seamlessly integrated alongside event-related information.

Other possible types of links might include predicates of perception, as well as possessive constructions, both of the predicative and adnominal kind.

22 — Semantically void predicates

Lastly, let us test the claim that the intransitive predicates commonly used for introductions tend to be semantically void in accordance with the idea of singling out new information from the rest of the narrative.

23 — Predicate semantics

[GRAPHIC: % of introductions by semantic predicate type in two of the Multi-CAST corpora]

Shown here are the relative proportions of various semantic predicate types used with new referents in two of the Multi-CAST corpora, Northern Kurdish and Mandarin. The rest are actively being worked on, but unfortunately weren't ready in time for this talk. Therefore, it is best to understand these observations as highly preliminary.

The leftmost type, effect, captures all predicates that bring about a change of state or location, and as such are chiefly transitive. Existentials are included among predicates expressing states-of-affairs, the fourth type from the left.

24 — Semantically void predicates

The intransitive predicates commonly involved in the introduction of new referents are not obviously semantically void, as claimed by Du Bois (1987).

State predicates, which include semantically empty items among others, make up between a third and a fifth of the predicates used with introductions in our pilot investigation. Motion predicates are slightly less common; change-of-state transitives are more common in Kurdish, less so in Mandarin. The latter also has a consistently higher rate of existential constructions in the Pear Story data we have discussed earlier. This highlights that there is a substantial degree of cross-linguistic variation in this respect, as well as a heavy impact of content differences. Keep in mind that these data are preliminary; a broader sample taken from different languages and multiple speakers will be necessary to make these observations more robust.

Even so, transitives are by and far the most frequent predicate type used with new information.

This includes verbs of creation and perception verbs, which introduce new information as stimuli, for example.

As we have also seen earlier with the Pear Stories, intransitives used with introductions are often motion events. We argue that their selection is not so much motivated by considerations of information management, but rather by the specific content-driven needs of the narrative, for instance where referents actually move in some direction when first entering the scene, in the Pear film often towards the camera. As noted earlier, intransitive introductions are also often semantically enriched by added adjuncts. All this runs counter to studies that have taken Pear

Story data as evidence for the cognitively challenging nature of introductions, as for instance in Kumagai (2006).

25 — No specialization

Apart from the overall high frequency of transitive object and oblique introductions, we find no obvious generalizations regarding the specialization of any role, constructions, or semantic predicate type for the introduction of new referents.

Notably, the differences between and among the Multi-CAST data and the Pear Stories suggest that in addition to language-specific conditions, narrative content plays a significant role in the way new information is introduced. Such a degree of variation is not what we should expect if introductions were indeed generally cognitively costly and syntax universally adapted to this challenge.

26 — Integration

Instead, interlocutors focus on conceptual content and states of affairs. New information is not given special billing, but is seamlessly integrated into the narrative, involving primarily referential choices and other decisions on the NP-level, such as definiteness marking.

One possible exception are major episodic breaks, which may require additional signalling, and hence are epiphenomenally associated with new referents, leading to confirmation bias.

27 — Role profiles

Introductions in A are simply avoided because of the convergence of humanness and topicality, as we have argued elsewhere. The S role is semantically broader and overall very frequent and hence gets an overall larger share of the introductions in a given text, but it is itself not specifically associated with new information. P, obliques, and other non-subject roles are comparatively less frequent; however, much larger proportions of their mentions are new as they are naturally associated with semantic roles that link to new information, for instance via perception or creation.

28 — Cognitive demands of introductions

Our results are not ultimately conclusive on whether or not introductions are cognitively more demanding; it is possible that referent introduction is a challenge for processing, but that this is not reflected in linguistic structure, at least not by way of deploying specific morphosyntactic constructions. Note also that given mentions outnumber introductions by a factor of ten.

An exception could be the role of transitive constructions, but here the solution is clearly not to separate new information, but to link it to the narrative. Other production-related discourse features could provide clues that result in implicit learning, see for instance the work of Jennifer Arnold (e.g. 2003).

29 — Cognitive demands of introductions

It is however not generally clear that and why introducing new referents is more demanding than the task of keeping track of and distinguishing between old referents. What is the metric for

meaningfully distinguishing between the demands of dealing with new vs. given information? After all, for the speaker, new information is not actually new, but is already kept in working memory prior to its verbalization.

A case could be made for the inverse perspective, that is, that it is in fact comparatively easier to deal with given information than with new. The notion that introductions are challenging might be then be the consequence of an undue focus on episode- or scene-setting contexts, where presentational constructions are indeed prominent. But this in turn can be argued to be an attention-directing device – "I'm about to recount a new episode, look at me! Wheee!" – rather than a coping mechanism for dealing with a new information. That is, the association of these contexts with new information is epiphenomenal, since they happen to often, but not exclusively, involve new referents.

30 — Summary: Referent introduction

In sum, our findings show that referent introduction is substantially less disruptive to syntax than has been claimed. In particular, we have found little evidence for, one, the separation of new information from general content advancement; two, the isolation of new information in intransitive constructions; and three, the association of new information with semantically vague verbs.

Additionally, the association of presentational constructions with new information may be an artefact of their prominent use at the beginning of texts and episodic breaks in narratives; their use is in fact not confined to new referents.

31 — Summary: Referent introduction

We therefore conclude that the introduction of referents is primarily driven by the demands of content, and that speakers seamlessly integrate new information into existing syntactic schemas on the fly, without recourse to using syntax that is specialized for the task.

Our findings also cast doubt on the fundamental claim that referent introductions are cognitively demanding, or at least more so than other aspects of discourse management.

Thank you kindly for listening to this talk. We look forward to answering your questions during the live session on Thursday, August 27, from 9:30 to 9:40 (GMT+2). The remainder of this recording briefly summarizes the corpus data and associated methodologies. Thank you again, and have a nice day.