Jocelyn Aznar and Frank Seifart*

Direct speech, silent pauses, speech verbs, and basic word order: a comparative corpus study of 12 languages

Abstract: This study investigates cross-linguistically the position of speech verbs and the occurrence of silent pauses relative to direct speech reports in narrative oral texts. We hypothesize that the position of speech verbs depends on the basic word order of verbs and their complements (VO vs. OV), and that the occurrence of speech verbs and pauses exhibit complementarity based on their shared function of marking the onset or end of reported speech. We use language documentation corpora from an areally balanced sample of 12 languages. We show that in most VO languages speech verbs exclusively precede speech reports, while in OV languages they may precede or follow them, or intervene, as middle verbs. Speakers of VO languages very often pause after speech reports, i.e. the position where there is no speech verb, while speakers of OV languages often pause before the speech report, and almost never between a speech report and a following ending verb. We conclude that basic word order substantially influences how speech reports are construed and prosodically phrased, in addition to asymmetries resulting from constraints on sequential ordering that are specific to complex units in general and to reported speech in particular.

Keywords: corpus-based typology; silent pause; direct speech; word order; speech verbs

1 Introduction

Speakers of every documented language can refer to other speech events within their discourse, regardless of whether these speech events may be thoughts, spoken

*Corresponding author: Frank Seifart [fʁaŋk ˈzaɪ.fɑːt], Structure et Dynamique des Langues (SeDyL), Paris Villejuif, France; and Humboldt-Universität zu Berlin, Berlin, Germany, E-mail: frank.seifart@cnrs.fr. https://orcid.org/0000-0001-9909-2088
Jocelyn Aznar [ʒos.ˈlɛ̃ az.ˈnaɣ], Universität Zürich, Zürich, Switzerland; Universität Bern, Bern, Switzerland; and Centre de Recherche et de Documentation sur l'Océanie (CREDO), Marseille, France, E-mail: contact@jocelynaznar.eu. https://orcid.org/0000-0003-0286-637X

words, or signs. This ability to report speech is central to the "reflexivity" (de Brabanter 2023) of human communication. It is well known that the grammatical means speakers can employ to represent speech events vary cross-linguistically, in terms of the availability of different subtypes of speech reporting and potential indexical shift, among others (Evans 2013). Much less studied, especially from a typological perspective, are prosodic characteristics of reported speech. One of these characteristics is silent pauses that occur at the edges of speech reports and help to delimit speech reports from the surrounding discourse. Example (1) from the Northwest Amazonian language Bora illustrates the occurrence of silent pauses, represented by brackets in the transcription with indications of pause duration in seconds and milliseconds, at two points of transition between a speech report and surrounding discourse. It also illustrates two occurrences the speech verb *neélle* 'she said'.

(1) Bora 0279_doreco_bora1263_mc_bora_ajyuwa[1]
  (0.2) tsáhaá neélle (1.8) íllure ihdyu o úllelle ámúhadívú ó wajtsɨ́hi neélle

tsáhaá	neé-lle	íllu-re	ihdyu	o	úlle-lle
no	say-F.SG	so-REST	like_this	1SG	walk-F.SG

ámúha-dí-vú	ó	wajtsɨ́-hi	neé-lle
2PL-ANIM-ALL	1SG	arrive-PRED	say-F.SG

  ' "No," she said, "I was only walking around and so I came to you," she said.'

The current study investigates typological patterns of reported speech regarding the relation between the position of speech verbs, the position of pauses, and the basic word order of a language. Our focus is on direct speech, as one type of speech reporting. Specifically, we address, firstly, whether, across languages, the position of speech verbs, like SAY, TELL, ASK, relative to speech reports correlates with the basic word order of a language, namely, whether speech reports pattern with verb complements and therefore tend to follow speech reports in OV languages and precede

---

1 Data sources are indicated by the annotation unit number, the language identification code, and the name of the text, as given in the corpora used here. See Table 1 for the bibliographical references to these corpora. The following abbreviations are used in morphological glossing, and these follow the Leipzig Glossing Rules (https://www.eva.mpg.de/lingua/pdf/Glossing-Rules.pdf): 1 – first person, 2 – second person, 3 – third person, ABL – ablative, ACC – accusative, ALL – allative, ANIM – animate, CLF – classifier, CON – conjunction, COP – copula, CVB – simultaneity, DEM – demonstrative, DIM – diminutive, DIR – directional, DL – dual, F – feminine, FUT – future, GEN – genitive, IC – initial change, INAN – inanimate, INDF – indefinite, IPFV – imperfective, IRR – irrealis, M – masculine, MED – medial, MNR – manner, NEG – negation, NMLZ – nominalization, NOM – nominative, O – object, PERF – perfective, PL – plural, POSS – possession, PRED – predicative, PST – past tense, QUOT – quotative, REC – recent, REM – remote, REP – repetitive, REST – restrictive, S – singular, SG – singular, TR – transitive, VOC – vocative.

them in VO languages. We then move on to examine the occurrence of silent pauses at the edges of speech reports. Here, we study whether silent pauses preferably occur at those edges of speech reports (before or after) where – in a given language – the speech verb does not occur, because in these positions they most effectively demarcate speech reports from surrounding speech. To fulfil this function, silent pauses should occur more often before speech reports in OV languages and more often after speech reports in VO languages. Our study thus contributes to establishing whether there are typological differences in prosodic phrasing of speech reports, and the relation of these differences to the basic word order of a language. More generally, our study also contributes to understanding the co-variation of prosody and syntax.

We approach these questions by studying speech reports in corpora of narrative texts from a world-wide and diverse sample of 12 languages. We apply a comparative-corpus approach that takes up the challenge of incorporating variability typical of spontaneous speech production documented in spoken corpora into cross-linguistic comparison for the sake of enhancing the ecological validity of typological generalizations (Barth et al. 2021; Levshina et al. 2023; Schnell and Schiborr 2022). The current study was carried out within the QUEST (Quality-ESTablished) project, in particular within the RefCo initiative (Aznar 2020; Aznar and Seifart 2020, 2022). This initiative aims to improve the quality of language documentation corpora in terms of annotation and metadata, and in particular to demonstrate the reusability of curated high-quality language documentation corpus data. The data presented here are among the outcomes of this initiative.

The approach taken here enables the current study to make novel contributions compared to three particularly influential previous seminal studies: Firstly, it adds corpus evidence and a world-wide coverage to Güldemann (2008), who studied direct speech in a sample of 39 African languages, based on information gleaned from grammatical descriptions. Secondly, it also adds corpus-based evidence, as well as evidence specific to speech reporting, to findings from Schmidtke-Bode and Diessel's (2017) cross-linguistic study on complementation. Finally, it adds cross-linguistic evidence to Malibert and Vanhove's (2015) seminal study on the interaction of word order and prosody in direct speech in four Afro-Asiatic languages.

The current study proceeds as follows: We start by situating our research within the existing literature on reported speech and presenting our research questions and hypotheses (Section 2). We then proceed to detail our data and methodology (Section 3). Then we present the results of our study (Section 4), followed by a discussion of these results (Section 5) and a brief conclusion (Section 6).

2 Theoretical background and research questions

In a classical definition, Voloshinov (1986: 115) proposed that "reported speech is speech within speech, utterance within utterance, and at the same time also speech about speech, utterance about utterance." His definition insists that reported speech is about intertwining a current text with another one. Along these lines, Tannen (1986), studying storytelling in Greece, argued that the term REPORTED SPEECH is misleading. She proposed the term CONSTRUCTED SPEECH instead to emphasize that reported speech is a (literary) performance by the current speaker. While we agree with her point, replacing the term REPORTED SPEECH by CONSTRUCTED SPEECH would stress one non-essential aspect of the phenomena, its performance aspect, while downplaying a core characteristic, the reference aspect. We thus prefer keeping the more traditional terminology of REPORTED SPEECH for referring to another speech event.

Much of previous research on speech reporting has focused on the distinction between "direct speech" and other forms of speech reporting, such as "indirect speech". Following Evans (2013: 68), we define direct speech here by three criteria: (i) it is presented as if it were original speech, (ii) it includes linguistic particularities of the original such as dialect, and (iii) deictically sensitive expressions such as demonstrative and tense inflection reflect the perspective of the original speaker. It is important to note that – even if direct speech is produced as if it had been already produced in a different context – it does not necessarily represent it accurately. Additionally, as Coulmas puts it, "the notion of verbatim rendition, that is, identity of form, seems to be culturally variable" (1986: 1).

Unlike direct speech, in indirect speech, speakers report another speech event while maintaining their own perspective. Indirect speech displays cross-linguistically variable combinations of deviations from "canonical direct speech" (Evans 2013), including different possibilities for indexical shifts in pronouns, tense shifts, and deviations from main clause syntax. To account for some of this variability, Aikhenvald (2008: 383) proposed the category of semi-direct speech that "[…] involves coreferentiality between the current speaker – rather than the author of the speech report – and a participant within the speech report", arguing that subtypes of reported speech are not clearly distinguished categories but focal points on a continuum. Another deviation from canonical direct speech are special sets of logophoric pronouns, best known from Niger-Congo or Afro-Asiatic languages, to reduce referential ambiguity in speech reports (Culy 1997; Hagège 1974; Pearson 2015). Nikitina and Bugaeva (2021) therefore argue that "logophoric speech" constructions constitute a different type of deviation from direct speech than "indirect speech" and thus also argue against a uni-dimensional typology of direct versus indirect speech, in line with Evans' (2013) canonical approach.

Importantly for the current study, direct speech, as opposed to indirect speech, appears to be cross-linguistically not only common, but it also appears to have relatively uniform characteristics across languages. These include, according to Evans (2013: 70) "independent, 'vivid' intonational contour in quoted material", "quote can be interrupted by quotational clause", and the fact that an "overt quotational clause [is] optional". These characteristics inform the typological comparison of the current study.

In addition to classical speech verbs, such as SAY, TELL, many languages have non-predicative morphemes called quotative markers that mark the beginning or the end of speech reports (Güldemann 2008: 122; Spronck 2012). While expressions such as English *be like* are sometimes also called quotatives (e.g., Blyth et al. 1990; Buchstaller and van Alphen 2012), in the current study, we label such expressions as speech verbs and use the term quotative only for grammatical morphemes that mark speech reports. Direct speech reports without any speech verb, or quotative marker, have been extensively documented by Jordanoska et al.'s (2022) and Güldemann's (2008) cross-linguistic studies. Lack of speech verbs appears to be particularly common in reported dialogues, between contributions by different participants of the dialogue, in a range of languages (e.g., Galucio 2024: 7–9; Jordanoska et al. 2022; Malibert and Vanhove 2015:8).

Prosodic aspects of speech reporting have been studied from a conversation analysis perspective for German and English (Couper-Kuhlen 1999; Günthner 1999; Klewitz and Couper-Kuhlen 1999) where a number of specific prosodic characteristics of speech reports have been identified. Among these, a combination of various prosodic cues help to mark stretches of speech as direct speech. Klewitz and Couper-Kuhlen (1999: 482), based on qualitative analyses of English conversational data, find that, "among the prosodic and paralinguistic devices used most frequently [for marking reported speech], are global pitch (register) and loudness shifts, global changes in speech rate and shifts to isochronous timing". In this account, "rhythmic pauses" preceding speech reports may help to mark the beginning of the speech report. On the other hand, pauses that interrupt the rhythmic structure may trigger additional pragmatic inferences regarding the reported speech. While analyses of speech rhythm and pragmatics are outside the scope of the current study, these findings highlight the importance of pauses for the construal of speech reports, especially pauses at the left edge of speech reports. Other prosodic aspects of direct speech that have been described in other languages include variable prosodic salience (Genetti 2011), as well as an increased speech rate and raising intonation (Demers 2009; Hanote 2015; Izre'el and Mettouchi 2015), and interactions of f0 and pauses (Leandri 1993; Michno 2021), and a greater overall pitch range (Cervone et al. 2015; Jansen et al. 2001).

The current study exclusively focuses on pauses at the boundary of direct speech and the surrounding discourse, which help to mark the beginning and end of speech reports. Based on a comparative study on direct speech in 39 African languages, Güldemann (2008: 219–224) recognized the importance of pauses for "reorientation", i.e. to mark the transition from surrounding discourse to a speech report or vice versa. He observed that pauses occur more often directly before speech reports than directly after speech reports (Güldemann 2008: 222), in addition to differences in the structure of speech verbs that precede versus those that follow speech reports. However, he did not relate these observations to word order and he fully acknowledged the tentative nature of his observations regarding pauses given that he did not have access to "objective" measures taken in audio-recorded spoken corpora but had to rely on impressionistic observations by language experts.

The systematic study of prosodic cues of speech reports, including pauses, from a typological perspective has been pioneered by Genetti's (2011) analysis of direct speech in Dolakha Newar, spoken in Nepal. Taking this approach a step further, Malibert and Vanhove (2015) apply the descriptive tools and analysis of Genetti (2011) to a typological comparison of spoken corpora of languages with different word orders. They studied prosodic boundaries at the edges of speech reports comparatively in corpora of Beja (Cushitic), Zaar (Chadic), Juba Arabic (Arabic based pidgin), and Modern Hebrew (Semitic). These four languages belong to different branches of Afro-Asiatic, but, crucially, they display different word orders. Based on their findings, Malibert and Vanhove tentatively propose that:

> "In SOV languages where the quotative verb follows the speech reports, their onset is systematically set off from the previous intonation unit, a clear prosodic cue, marking the beginning of the speech report. In SVO languages it is the end of the speech report which is set off from the next IU [intonation unit]." (Malibert and Vanhove 2015: 61)

This is, to the best of our knowledge, the first observation regarding a typological correlation between word order and prosodic boundary marking at the edges of speech reports. It provides a major motivation for the current study, which sets out to expand and test this proposal on a broader sample of languages. We first break this observation up into two testable hypotheses. Our first hypothesis relates to the position of speech verbs, building on previous research on the position of complement clauses (Schmidtke-Bode and Diessel 2017) and on reported speech in African languages (Güldemann 2008):

Hypothesis 1: Speech reports behave like complex verb complements regarding their position relative to speech verbs, i.e. speech reports tend to follow speech verbs in VO languages and mostly precede them in OV languages, although some will follow speech verbs in OV languages, too.

Note that we are not claiming that speech reports are syntactically identical to complement clauses in the languages studied here – there is plenty of evidence of the syntactic differences between the two (e.g., McGregor 1994; Munro 1982; Spronck and Nikitina 2019), also regarding permitted word orders with respect to complement taking predicates and speech verbs, respectively (Longacre 2007: 387–388). Rather, our hypothesis is based on a basic parallelism between the two constructions in that they are both headed by a verb and involve a potentially complex and potentially clause-like dependent element. The ordering pattern expressed in our hypothesis results from two competing ordering motivations: Consistent head-depended ordering on the one hand, and the tendency to place relatively heavy constituents later in a clause, including complex noun phrases, relative clauses, and complement clauses (e.g. Behagel 1909; Dik 1997: 99–416; Dryer 1980; Hawkins 1994), as discussed in detail regarding reported speech by Güldemann (2008: 210–219). The latter principle explains why we expect speech reports to sometimes follow speech verbs in OV languages even though this introduces an inconsistency in terms of head-dependent ordering.

Our second hypothesis relates to the occurrence of pauses before and after speech reports, and we again further break this hypothesis up into two hypotheses:

Hypothesis 2.1: Speech verbs and pauses display complementarity, i.e., silent pauses will occur at those edges of speech reports where speech verbs are absent.

Hypothesis 2.2: Silent pauses will occur more often before than after speech reports in OV languages and more often after than before speech reports in VO languages.

Underlying Hypotheses 2.1 and 2.2 is the assumption that silent pauses are a reliable cue for the identification of prosodic boundaries, which help to demarcate the limits between speech reports and the surrounding speech. We break this hypothesis up into two because we are interested in finding out to what extent pause occurrence is directly linked to the absence of a speech verb (Hypothesis 2.1) and to what extent pause positions are also determined by basic word order, possibly independently of the actual presence or absence of a speech verb (Hypothesis 2.2). We focus on the presence versus absence of pauses in the current study, but we include some results on pause duration, with expectations for pause durations analogous to those for pause probability. Regarding both speech verb position and pauses, we focus here on introductory and ending verbs, but we also report results on middle verbs.

We are aware that prosodic boundary marking involves language-specific combinations of several cues, including pitch reset and final lengthening, besides pauses (Himmelmann and Ladd 2008). However, silent pauses appear to be cross-linguistically common enough as cues for prosodic boundaries (Peck et al. 2021) to

justify a study which focuses on silent pauses alone. This choice also keeps our measurements across languages comparable, especially in the absence full-fledged prosodic analyses and annotations of corpora on 12 languages, which would have been beyond the scope of the current study. We are also aware that, besides functioning as boundary markers, speech pauses respond to speakers' need for planning and breathing so that more pausing is expected after and before longer stretches of speech (Grosjean and Collins 1979; Henderson et al. 1965; Torreira et al. 2015; Włodarczak and Heldner 2017). We take this into account in our analyses, although, for technical reasons, only for pauses before speech reports. In as much as speech reports are clause-like syntactic units, we expect pause probabilities at boundaries of speech reports to be in the same range as those that have been observed at corresponding syntactic boundaries. Specifically, Peck and Becker (2024) observed in corpora from a diverse set of seven languages pause probabilities of between 29 % and 57 % at dependent clause boundaries and probabilities between 50 % and just over 75 % at main clause boundaries. Hypothesis 2.2 thus aims at finding meaningful differences within these ranges depending on the presence or absence of a speech verb and the word-order type of the languages.

3 Data and methods

3.1 Data: corpora on 12 diverse languages

The 12 corpora used in the current study were extracted from the set of 51 time-aligned corpora included in the DoReCo collection (version 1.0, Seifart et al. 2022). Each of these contains on average around 10,000 words of mostly narrative speech that has been collected through on-site fieldwork on often endangered languages, and transcribed, translated, and annotated by experts on these languages. For 38 languages, this annotation also includes morpheme segmentation and interlinear morpheme glosses. From this set we selected 12 languages for the current study (Table 1). This selection was based on the following criteria: (i) equal representation of VO and OV languages as well as inclusion of some languages with no dominant word order (also called non canonical word order), and (ii) world-wide genealogical and areal diversity. While this language sample is still limited, it is three times as large as that of Malibert and Vanhove (2015) and represents the most comprehensive dataset available for this type of labor-intensive manual annotation of direct speech at present. Information in Table 1 includes the Glottocode language identification code, which points to entries in the Glottolog (Hammarström et al. 2022), from where information on geographic area and on top-level genealogical language family was taken. Information on basic word order (BWO) was taken from Dryer (2013), where

available, or from published descriptions of the language. We additionally provide information on the "genus" a language belongs to, i.e. on genealogical units of comparable time depths (roughly 4,000 years) that are commonly used in sampling for typological studies (Dryer 1989). This information was taken from the *World Atlas of Language Structures* (WALS) (Dryer and Haspelmath 2013).

Four languages from the Americas are represented in our sample, three South American (Bora, Movima, and Mojeño Trinitario) and one North American language (Arapaho). The latter is a language of the Algonquian branch of the Algic family, spoken by around 250 speakers according to Cowell and Moss (2008). The language is highly polysynthetic and does not have a dominant word order regarding the placement of V and O. The Bora language is spoken by around 3,000 speakers in the Peruvian and Colombian Amazon and belongs to the small Boran family. Bora is a tonal language with OV word order, where SOV and OSV are equally common. Since we are interested in the order of objects, not subjects, relative to verbs, we group it with other (S)OV languages here. The language isolate Movima is spoken in Bolivia by

Table 1: Language sample used in the current study.

	Language	Glotto-code	BWO	Area	Family	Genus	Corpus reference
1.	Arapaho	arap1274	No dominant	North America	Algic	Algonquian	Cowell (2022)
2.	Movima	movi1243	No dominant	South America	(Isolate)	(Isolate)	Haude (2022)
3.	Beja	beja1238	SOV	Africa	Afro-Asiatic	Beja	Vanhove (2022)
4.	Bora	bora1263	SOV	South America	Boran	Boran	Seifart (2022)
5.	Dolgan	dolg1241	SOV	Eurasia	Turkic	Turkic	Däbritz et al. (2022)
6.	Sanzhi Dargwha	sanz1248	SOV	Eurasia	Nakh-Daghestanian	Dargwic	Forker and Schiborr (2022)
7.	Savosavo	savo1255	SOV	Papunesia	(Isolate)	(Isolate)	Wegener (2022)
8.	Fanbyak	orko1234	SVO	Papunesia	Austronesian	Oceanic	Franjieh (2022)
9.	Mojeño Trinitario	trin1278	SVO	South America	Arawakan	Bolivia-Parana	Rose (2022)
10.	Nisvai	nisv1234	SVO	Papunesia	Austronesian	Oceanic	Aznar (2022)
11.	Nǁng	nngg1234	SVO	Africa	Tuu	Tuu	Güldemann et al. (2022)
12.	Ruuli	ruul1235	SVO	Africa	Atlantic-Congo	Bantu	Witzlack-Makarevich et al. (2022)

about 1,400 speakers. Word order in Movima follows complex syntactic rules, with no dominant order for V and O (Haude 2006). Finally, we use a corpus of the Arawakan language Mojeño Trinitario, also spoken in Bolivia. This language has been described as having SVO word order (Rose 2014: 89). For reasons of space, we refer to this language as "Mojeño" in the remainder of the current study.

Among the three African languages in our data set, Beja is an Afroasiatic language from the Cushitic branch spoken by one to two million speakers in Egypt, Sudan, and Eritrea and it has SOV word order (Vanhove 2017). The corpus used for the current study is a subset of the corpus used in Malibert and Vanhove's (2015) study. N‖ng, from the Tuu family (formerly classified as "Khoisan"), is a critically endangered language spoken in Namibia. It has been described as an SVO language. Ruuli is a Bantu language spoken in Uganda by around 160,000 speakers and it is also an SVO language (Sørensen and Witzlack-Makarevich 2020).

There are three corpora from Oceania (called "Papunesia" in Glottolog) in our study, two Austronesian languages, Fanbyak and Nisvai, and the isolate Savosavo. Both Fanbyak and Nisvai are spoken in Vanuatu, the first one on Ambrym, and the second one in South-East Malekula, by around 200 native speakers each, and they both belong to the Central Vanuatu branch of Oceanic. They are both SVO languages. The Nisvai corpus used for this study is part of a larger corpus that was created for a study of narrative practices within the Nisvai community (Aznar 2019). Savosavo is an isolate language spoken by about 3,000 speakers on Savo Island in the Solomon Islands, and – unlike Nisvai and Fanbyak – it is an SOV language (Wegener 2012).

Two languages from Eurasia are represented in our sample: Dolgan is a Turkic language spoken on the Taymyr Peninsula in Siberia by around 1,000 speakers. Like other Turkic languages, its word order is SOV (Stapert 2013: 246). Sanzhi Dargwha is a Nakh-Daghestanian language spoken in the Northern Caucasus by around 250 speakers, and its basic word order is also SOV (Forker 2019). We refer to this language as "Sanzhi" here.

We selected between one and nine texts from each of these 12 DoReCo corpora that we further annotated for the purpose of the current study, as summarized in Table 2. This selection strikes a balance between keeping the amount of manual annotation feasible, on the one hand, and representing speakers of different sexes and age groups where possible. Data processing was facilitated by consistency across DoReCo data regarding the naming and structure of annotation tiers, such as transcription, word-level segmentation, interlinear glosses, as well as the time-alignment of these annotations.

Despite our efforts to produce a balanced data set, Table 2 shows various biases and imbalances regarding the number of texts, their lengths, and the characteristics of the speakers. For instance, Bora and Savosavo represent almost 40 % of the total corpus in terms of interpausal units or recording duration. Regarding the sex of the

Table 2: Data set used in the current study.

Language	Speaker sex	Speaker age (av.)	Length (minutes)	Number of texts	Number of interpausal units	
Arapaho	m	70	16.7	1	487	487
Beja	f	40	8.3	1	221	941
	m	40	27.2	8	720	
Bora	f	55.5	19.3	2	409	1,802
	m	55.75	56.8	2	1,393	
Dolgan	m	72	24.1	1	519	519
Fanbyak	f	54	5.6	1	145	483
	m	61	11.0	2	338	
Mojeño Trinitario	m	60	12.0	1	300	300
Movima	f	70	5.8	1	163	757
	m	54	25.2	1	594	
Nisvai	f	60	7.6	1	146	934
	m	44	43.0	5	788	
Nllng	f	71.5	8.0	2	287	535
Ruuli	f	60	7.3	2	248	248
Sanzhi	m	60	10.3	2	218	218
Savosavo	f	25	7.2	1	191	1,420
	m	55	42.9	2	1,229	
Total	f	59.5	69.2	11	1,810	8,396
	m	50.24	269.2	25	6,586	

speakers, there is also a discrepancy: 25 texts were produced by men and only 11 by women, and this disparity is even more important when we consider the time ratio between male and female speakers, which is almost four times more for male speakers. Another bias is that older speakers are overrepresented in our data: the average age of speakers in our data set is 53.1 years. These biases can be attributed to the fact that most DoReCo data stems from language documentation projects that focus on the documentation of traditional narrative practices, which are more often produced by elders of the communities, and often by men.

DoReCo corpora provide audio files and corresponding annotation files in ELAN format (ELAN Developers 2021), among others. We used these ELAN annotation files

as input for building a multilingual corpus with comparable units out of the time-aligned corpora. DoReCo data also already included annotation of silent pauses, as a result of automatic identification of silent pauses during automatic phone-level time alignment with subsequent manual corrections at the level of word start and end times, which paid particular attention to the identification of silent pauses (Paschen et al. 2020). These pauses define interpausal units (IPUs) in our data (see Table 2). To further process these data, a Python module was developed following the rationale laid out in Aznar (2020), in order to compile a single, cross-linguistic corpus according to our study's requirements, working with interpausal units. Furthermore, a Jupyter notebook (Kluyver et al. 2016) was written and combined with Poetry[2] to document and ease the reuse of our data processing methods. These tools also aided in producing the results presented in this article, as well as the visualizations that helped to interpret the results.[3]

3.2 Annotation conventions, data coding, and analyses

We developed a dedicated annotation scheme (Table 3) for testing our hypotheses and applied it across corpora. The annotation tags delimitate and describe direct speech reports and three positions of speech verbs: introductory, middle, and ending verbs. Tags for speech verbs in fact extend over entire verb phrases, potentially including its arguments, TAM marking, etc. Our annotation scheme also includes tags for interjections. Even though their analysis is outside the scope of the current study, we report on their occurrence, along with that of pauses, in Section 4.1. Following Norrick's proposal (2014) to consider phrases such as *Oh my god* as interjectional, we annotated not only interjectional morphemes as interjections but, whenever relevant, the whole phrase associated with it. We also annotated appellatives as interjections. These are often vocative kinship terms, as in *"Grandfather," he said, "come here".* Recall that silent pauses were already annotated in the DoReCo data we used as input.

Our annotation of direct speech is framed within the concept of direct speech sequence, which is a larger unit that contains all relevant elements surrounding speech reports, including potentially more than one speech verb. The concept of sequence is inspired by the work of Adam (2011) on text sequences as discourse units associated with linguistic cues. A direct speech sequence is characterized by containing one speech report, unless a middle verb occurs, in which case it contains two

---

2 See https://python-poetry.org for more information about Poetry, a tool for packaging and managing dependencies for Python.
3 These materials can be found at https://doi.org/10.5281/zenodo.14753908.

Table 3: Annotation conventions.

Annotation tag	Meaning
<<begin_sequenceDS>>	Beginning of a direct speech sequence
<<end_sequenceDS>>	End of a direct speech sequence
{{	Beginning of direct speech report
}}	End of direct speech report
[[iv:verb]]	Introductory verb (phrase) occurring before a speech report, for example *and* [[iv:*he said*]] {{*Hey, how are you, Bill?*}}
[[ev:verb]]	Ending verb (phrase) occurring after a speech, report for example {{[[inter:-*Hey*]], *how are you, Bill?*}}, [[ev:*he said*]]
[[mv:verb]]	Middle speech verb (phrase) occurring in between two direct speech reports, but belonging to both the preceding and the following speech report
[[qu:quotative]]	Quotatives, that is a non-predicative morpheme marking the beginning or the end of direct speech report
[[inter:interjection]]	Interjection phrase or appellative phrase (like vocative, kinship term or pronoun).

speech reports. It can, and often does, extend over multiple interpausal units. In our definition of a direct speech sequence, a sequence cannot contain the same type of speech verb twice: a second introductory verb initiates a new direct speech sequence. On the other hand, a direct speech sequence containing an introductory verb can be followed by another direct speech sequence containing only a middle verb. While according to the formal criteria, both could be within the same sequence, we relied on the content within the narrative for indications that the two speech reports are different events, e.g. when they report speech by two different speakers in dialogical speech, and then we annotate these as two separate sequences.

Each of the corpora in the current study had already been manually transcribed, translated, and morphologically annotated by the teams of authors that created them (see Table 1). We manually added annotations for the current study on direct speech, as given in Table 3, to a copy of the transcription tier. This annotation work was done by the first author for 11 of the corpora used here, and by the second author for the Bora corpus. Our identification of direct speech directly followed the language experts' use of quotation marks in free translations (and in some cases also in the orthographic transcriptions). This concerns the start- and endpoints of speech reports as well as potential borderline cases such as *Life is short*, which could be hard to classify as either direct or indirect speech (although such cases only very rarely occurred in our data). We also found that in the narrative texts we used (perhaps unlike in conversation), the context almost always makes it clear if speech is attributed to a character of the narrative or not. Where in doubt, we consulted the

language experts that created the corpus. We also made sure to only include instances of speech reports that fully conform to the definition of direct speech used in the current study (see Section 2). Manually annotating data ensured that the authors gained detailed insights into how direct speech is expressed in each of the languages, which facilitated a reliable interpretation of the results of the quantitative analyses.

Regarding the annotation of speech verbs, we only included speech verbs that are actually syntactically linked to a speech report. We thus excluded cases like *mwēwar* 'he spoke (to them)' in example (2) from Dolgan, which describes the speaking event in general. We did include the verb *mwigile* 'he said' (in bold), which actually introduces a speech report.

(2)   Dolgan: independent speech verb and introductory verb
        0072_doreco_SJ1_9_50 – 52
        (1.4) mwēwar vane gēlalō <<begin_sequenceDS>>[[iv:mwigile]] (1.4){{a sowe nge (0.4)
        lobungbung gola ru mwerame   lolēn   mwerame   lolne   go   abwimiin
        wē}}<<end_sequenceDS>> (1.0)

mwe=war	van-e	gēlalō	**mwi=gile**
3sg.rec.pst=speak	go=tr	3dl	**3sg.rec.pst=say**
a	sowe	nge	lobungbung	gola	ru	mwe=rame
con	?	3sg	child	dem.sg.med	stay	3sg.rec.pst=think
lolēn	mwerame	go	mwe=rame	lol-ne	abwi=min
?	?	?	3sg.rec.pst=think	insides=tr	3sg.irr=drink
wye
water

        'He spoke to the two of them, he said "What's this? That child is wanting, he
        wants to drink water".'

Example (2) also illustrates the first type of speech verbs, namely introductory verbs. Example (3), from Beja, illustrates an ending verb. In example (4), from Arapaho, the verb *nihʔiit* 'PST-said-3.S' is a middle verb in between two speech reports. Example (5), from Bora, illustrates an introductory, a middle, and a final verb within one direct speech sequence.

(3)   Beja: Direct speech sequence with an ending verb
        0077_doreco_BEJ_MV_NARR_08_drunkard_BEJ_MV_NARR_08_drunkard_125
        (0.77) {{na:na:d=da jʔ-a:-w-wa flan}} [[ev:e-ndi=ho]] (0.43)

na:na:d=da	jʔ-a:-w-wa	flan
what=dir	come-cvb.mnr=indf.m.acc=cop.2sg.m	so

        e-ndi=ho
        3sg.m-say.ipfv=when
        'When he says, "So, why have you come?" '

(4)   Arapaho: Direct speech sequence with a middle verb
        78a.eaf:0178_doreco_78a_78a.193
        (0.5) {{Wohei}}, (0.3) [[mv:nih'iit]], (1.9) {{woow (0.4) ni'P (0.2) nii'oo'.}} (0.7)

Wohei	**nih-ʔii-t**	woow	ni'	niiʔooʔ
okay	PST-**said-3.S**	now.PERF	good/well	IC.good

        ' "Okay," he said, "now it's okay." '

(5)   Bora 0109_doreco_bora1263_mc_bora_ajyuwa Bora
        (1.5) ane [[iv:diille íeeválleke neélle]] {{muúllej}} [[mv:neélle]] {{kiátúikyé u
        pájtyeíllejí͈́ivari}} [[ev:neélle]] (1.5)

aa-ne	**di-lle**	i-eeva-lle-ke		**nee-lle**
CON-INAN	**3-F.SG**	**3-be_pregnant-F.SG-ACC**		**say-F.SG**
muúlle-j	**nee-lle**	kiá-tu-iíkye	u	pajtye-i-lle-jí͈́ivari
sister-VOC	**say-F.SG**	where-ABL-yet	2SG	cross-FUT-F.SG-NEG
nee-lle				
say-F.SG				

        'And she said to the pregnant one, "Sister," she said, "Where could you cross?,"
        she said.'

For the definition of pauses in the context of direct speech production, we only
included silent pauses, and disregarded filled pauses. The reason for this is that filled
pauses appear to be alternatives for silent pauses only in their function as hesitation
markers, rather than as markers of prosodic boundaries, which is what we are
interested in here. Recall that DoReCo's time-alignment procedure involved manual
corrections specifically of start and end times of silent pauses. Therefore we did not
exclude any silent pauses, e.g. pauses shorter than any given minimal length.

Regarding the analyses of the data, we combine qualitative analyses with
descriptive statistics such as proportions of different verb types in languages with
different word orders. To estimate the effects of the different factors on the proba-
bility of silent pauses before and after speech reports, we fitted two Bayesian
Generalized Linear Mixed Models with logistic (Bernouilli) response, using PyMC 5
(Abril-Pla et al. 2023), one predicting the presence or absence of a pause before the
speech report, and another predicting the presence or absence of a pause after. As
fixed effects, we included the presence or absence of each of the three types of speech
verbs and the basic word order, as well as speaker sex and speaker age, standardized
as a z-score, to control for variation according to these two factors (e.g., Jacewicz et al.
2009); for the pause-before model only, we also incorporated the distance between
the onset of the speech report and the preceding last pause. Because of the nature of
our annotation, we were not able to measure the distance to a following pause, to
control for duration (or length) of the following phrase. Languages and speakers
were treated as random intercepts to capture variability across both groups. The

models include interactions between word order and the respective speech verb (introductory or ending) to determine whether languages with different word orders exhibit distinct pause practices. For the interpretation of results, the coefficients are estimated on a logit scale; each coefficient represents a change in the log odds of a pause occurring per unit change in the predictor. Therefore all the predictor effects are to be interpreted relative to the intercept level. Finally, we also followed the practices of Engelmann et al. (2019), Kumarage et al. (2022), and Donnelly et al. (2024) to distinguish between weak evidence (for results within the 85 % Highest Density Interval [HDI]) and strong evidence (results within the 95 % HDI). Results under the threshold of 85 % were treated as providing no evidence.

4 Results

4.1 Variability in the structure of speech reports

Before presenting results that directly address our hypotheses (Sections 4.2–4.3) and additional results regarding middle verbs (Section 4.4), we provide in this section some general observations regarding the cross-linguistic variability in speech verb usage and the structure of direct speech sequences. These observations are summarized in Table 4.

Table 4: Basic statistics on direct speech sequences, speech reports, speech verbs, and interjections.

Language	BWO	Direct speech sequences	Speech reports	Sequences without speech verbs	Sequences with interjections
Arapaho	No dom.	126	134	58.7 %	33.3 %
Movima	No dom.	48	55	47.9 %	45.8 %
Beja	S**OV**	156	169	15.4 %	27.6 %
Bora	S**OV**	160	231	10.6 %	63.1 %
Dolgan	S**OV**	135	179	16.3 %	58.5 %
Sanzhi	S**OV**	38	52	15.8 %	50 %
Savosavo	S**OV**	115	116	10.4 %	46.1 %
Fanbyak	S**VO**	37	37	13.5 %	43.2 %
Mojeño	S**VO**	46	49	52.2 %	28.3 %
Nisvai	S**VO**	162	162	19.8 %	80.2 %
Nǁng	S**VO**	53	53	54.7 %	52.8 %
Ruuli	S**VO**	37	37	37.8 %	50.8 %
Total/average		1,113	1,274	25.3 %	50.8 %

In Table 4, observe firstly that the number of speech reports exceeds the number of direct speech sequences in our data in many languages (although not in most SVO languages). This is due to the use of "middle verbs" in these languages, which break up direct speech sequences into various speech reports. Secondly, observe that direct speech may occur without any overt speech verb in all languages of our sample, consistent with Jordanoska et al.'s (2022) findings. The absence of a speech verb is in fact more common than its presence in four out of the 12 languages studied here, including both languages with no dominant order and Mojeño, which behaves like a language with no dominant word order in many respects (see Sections 4.2–4.4). Thirdly, we can also see that the use of interjections in speech reports is recurrent across the languages studied here, again to varying degrees, in line with Gülde-mann's (2008: 40) remark that "quote-initial expressive items like interjections, exclamations, vocatives, etc. are quite effective devices that the reporter can use for DRD-marking [Direct Reported Discourse]".

Another important observation is that the phrases headed by introductory verbs tend to be more complex compared to those headed by ending verbs. In example (6), from Savosavo, direct speech is distributed over three interpausal units and is associated with two speech verbs: the introductory verb *savu-li-ghu(e)* 'say/tell-3SG.M.O-NMLZ' and the ending verb *tei* 'want.to.do/say/be.like.this' (another instance of that verb, *tei-ghue* is used with the meaning 'want.to.do' within the speech report in this example). The phrase headed by the introductory verb *savu-li-ghu* does not only introduce the speech report here, but also specifies the participants involved in the situation, namely the third person speaker, through *tulola-lo*: (then-3SG.M.NOM), and the recipient, *mama* 'mother'. On the other hand, the phrase headed by the ending verb *tei* is much less complex, and shorter.

(6) Savosavo 0052_doreco_doreco_savo1255_ap_cs_saraputu
(0.9) Manamanali tulolalo <<begin_sequenceDS>>[[iv:mama kaka savulighu(e):]]
{{[[[inter:Mama]] (1.0) aiva kama (0.8) bo teighue}}[[ev: lona tei]]
<<end_sequenceDS>>(1.7)

Manamana-li	tulola-lo	**mama**	**k-aka**
prepare-3SG.M.O	then-3SG.M.NOM	**mother**	**3SG.F.O-with/to**
savu-li-ghu(e):	Mama	aiva	kama	bo
say/tell-3SG.M.O-NMLZ	mother	1SG.GEN	already	go
tei-ghue		lo=na	**tei**.
want.to.do/say/be.like.this-NMLZ		3SG.M.NOM=NOM	**want.to.do/say/be.like.this**

'He prepared it and then he said to his mother, "Mama, I'm already about to go," he said.'

Example (7) from Arapaho also illustrates this difference. The phrase headed by the first instance the verb *'iit* 'to say' explicitly mentions the character who produced the

speech report *howoh'oe* 'wait!'. This is particularly important in Arapaho narratives, where there are many instances where the characters are not overtly expressed.

(7)   Arapaho 0007_doreco_78a_78a.010
Noh, (0.43) <<begin_sequenceDS>>[[iv:nih-'iit nehe' hiisiis]], (0.16)
{{howoh'oe,}} (0.15) [[ev:nih-'iit]].<<end_sequenceDS>> (0.8)

Noh,	**nih-'iit**	**nehe'**	**hiisiis**,	howoh'oe,	**nih-'iit**.
and	PST-**said**	**this**	**sun**	wait!	PST-**said**

'And the sun said, "Wait!", he said.'

Example (8) from Bora contains another example of an introductory verb that heads a complex phrase (see example (5) for yet another example).

(8)   Bora 0015_doreco_bora1263_mc_bora_meenujkatsi
(0.8) <<begin_sequenceDS>> aanevápe [[iv:diibye tsaapi méwake nééhií]] (0.8)
{{aatye mahájkímú taabámú taúmeípíwu ícyahíjcyáhi}}<<end_sequenceDS>>

aa-ne-va-pe	**di-be**	**tsa-pi**	**mewa-ke**
CON-INAN-QUOT-REM	**3-M.SG**	**one-CLF.man**	**wife-ACC**

nee-hi	aátye	me-hajki-mu	taába-mu
say-PRED	those	1PL.POSS-relative-PL	wife-PL

táúmei-pi-wuu	ijcya-híjcya-hi
ask_for-in_excess-DIM	be-REP-PRED

'And then the man said to his wife: "These wives of our relatives keep begging excessively."'

We thus observe in all three cases that the introductory verb is used to (re-)specify certain aspects of the narrative situation, such as who the main characters are. Some languages, like Savosavo, in fact use different verb stems as introductory versus ending verbs (example 6).

4.2 Speech verb position and basic word order

In this section, we address our first hypothesis, namely that speech reports pattern with verb complements in terms of word order, i.e. they tend to follow speech verbs in VO language and mostly precede, but sometimes follow them in OV languages. Figure 1 summarizes the proportions of verbs occurring in different positions relative to the speech report found in our data, based on the counts of direct speech sequences and speech reports given in Table 4. The upper part of Figure 1 visualizes the proportions of introductory and final speech verbs (blue vs. green bars), as well as middle verbs (orange bars) in each language. The lower part reports on the
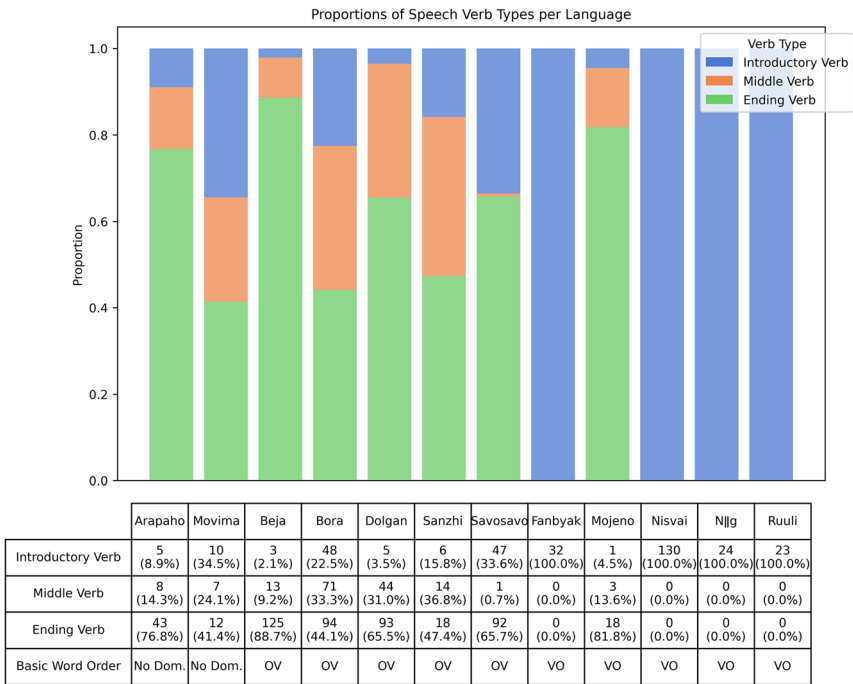
Figure 1: Occurrence of speech verbs in different positions with respect to speech reports. See Table 4 for the total numbers of direct speech sequences.

number of speech verbs observed and the percentage it represents per language and among the sequences with at least one speech verb.

Figure 1 shows that there is a clear relationship between the basic word order of a language and the position of speech verbs: In all five OV languages in our sample, speech verbs more often follow than precede speech reports, consistent with the basic order of verbs and their complements, but may follow them, consistent with the "heavy last" principle. In four out of five of the VO languages in our sample, speech verbs exclusively precede speech reports, i.e. these four languages only use introductory speech verbs, situated before the speech report (blue bars). Overall, results from nine out of ten languages that do have a basic word order are thus consistent with our hypothesis. Mojeño, however, is a stark exception in that there are almost no introductory speech verbs, and the majority are ending verbs. We confirmed the basic order of V and O in that language by analyzing their order in the first 71 instances of transitive non-speech verbs with nominal objects in the data we used here, and found that 68 of them had VO order.

We thus observe that the position of speech verbs in VO languages is practically fixed in the initial position, with the exception of Mojeño, while in other languages, it is much more variable: While all OV languages use more ending verbs (green bars) than introductory verbs (blue bars), they all also use introductory verbs and middle verbs (orange bars), separately or together, as we will discuss in Section 4.4. Speakers of Bora, for instance, use as many ending verbs as middle verbs, in addition to rare cases of introductory verbs. This flexibility is consistent with the fact that the order of objects relative to the verb is relatively flexible in Bora. In our Beja data, on the other hand, the great majority of speech verbs are ending verbs, consistent with Malibert and Vanhove's (2015) findings, and also consistent with the fairly strict OV order they report for Beja.

Finally, we observe that the two languages with no dominant order of verbs and objects, Arapaho and Movima, pattern with OV languages, rather than VO languages, both in terms of the preference for ending verbs, and in terms of using middle verbs. Regarding middle verbs, we observe that these only occur in languages which also have ending verbs,[4] and these two verb types are a feature exclusively of OV languages and languages with no dominant order, again with the exception of Mojeño. Finally, as can be seen from the percentages reported at the bottom of Figure 1, and confirming our observations from Table 4, above, languages with no dominant order, and – again – Mojeño, use overall considerably fewer speech verbs than other languages.

4.3 Pauses at the edges of speech reports: complementarity and basic word order

We hypothesized that silent pauses occur at those edges of speech reports where the speech verb does not occur, as a cue for identifying transitions to or from direct speech within a narrative. In this section, we address this by examining (i) the complementarity of pauses and speech verbs, i.e. how pause occurrences relate to occurrences of speech verbs in a given context (Hypothesis 2.1), and (ii) how the occurrence of a silent pause before and after speech reports relates to the basic word order of a language (Hypothesis 2.2). We first present our main results addressing these two hypotheses regarding the presence versus absence of pauses, and then complementary results regarding pause durations.

---

4 We are aware of exceptions to this in Nisvai data that was not included in DoReCo and consequently also not included in the current study: In a larger Nisvai data set (Aznar and Gala 2020) there is at least one occurrence of a middle verb, but no ending verbs.

Figures 2 and 3 present the results from our Bayesian generalized linear mixed models for estimating the effects of speech verb presence, as well as those of other factors, on the probability of a silent pause before and after speech reports.

For pauses before speech reports, Figure 2 shows an intercept mean of 1.951, which corresponds to a probability of 88 %, for the presence of a pause before speech reports under the model's baseline scenario, which is defined as OV word order, absence of any speech verbs, average speaker age, and baseline sex (defined here as male). When an introductory verb is present, compared to no speech verb in the same baseline scenario, the probability of a pause before speech reports is strongly reduced: the result for "Intro Verb" has a mean of -2.040, which corresponds to a probability of 48 % for a pause to occur before speech reports. Further results reported in Figure 2 put this result into a wider perspective. Firstly, the presence of a speech verb in another position, middle or ending, does not have a credible effect on the production of a pause before the speech report. Regarding the effects associated with speaker characteristics, we observed a credible, positive age effect (mean $\approx$ 0.506), indicating that older speakers are more prone to pause at the onset boundary. Recall from Table 2 that speakers in the corpus are relatively old, 53.1 years on average, and the youngest speaker 25 years old. The effect of female sex on pausing is less clear: The mean of the effect is slightly negative (mean $\approx$ -0.323), but its Highest Density Interval (HDI) overlaps with zero, reflecting uncertainty or weaker stability of this effect. In Figure 2, we also report on the effect of the distance between the speech report and the preceding pause ("Distance to Last Pause" in Figure 2). This distance has been calculated by counting the number of transcribed characters between the beginning of the speech report and the preceding pause – regardless of whether there is a pause directly preceding the speech report, which is what we are predicting. Results indicate that the longer this distance, the more likely the presence of a pause before a speech report. This shows that, as expected, there is an increasing need for pausing the longer the distance since the last pause, for reasons of prosodic phrasing, and maybe also the physiological need for breathing. As we investigate this effect in a single model together with the effect of speech verb presence, we are able to show here that the presence of a speech verb independently and additionally has an effect on pause probabilities.

Regarding pauses after speech reports, Figure 3 shows a positive effect of 2.47 on the logit scale for the intercept, which means that in the baseline scenario (OV language, no speech verb, average-age male speaker), there is a probability of 92 % for a pause to occur after a speech report. Compared to the baseline scenario, the
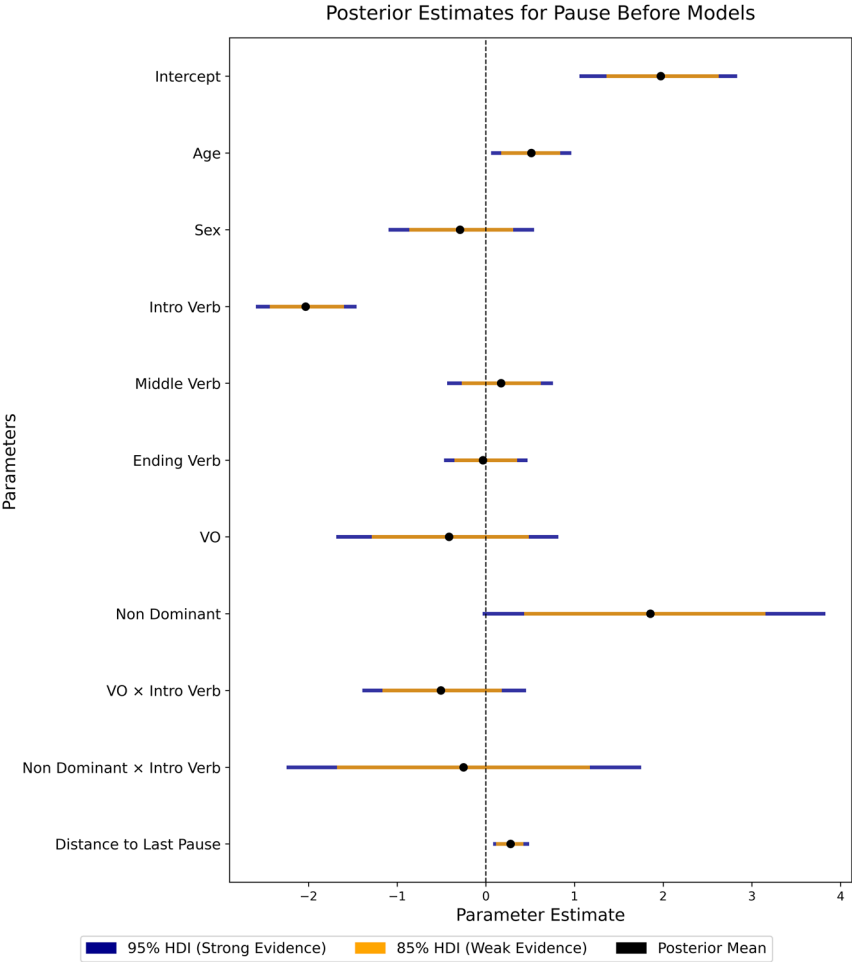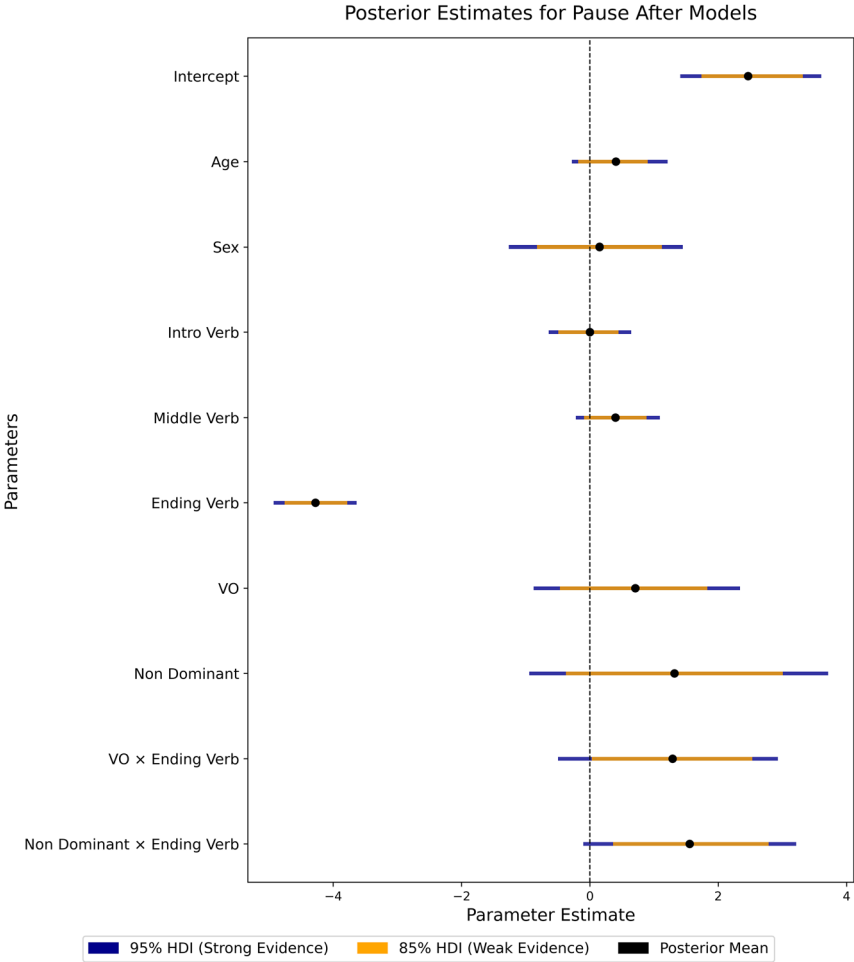
Figure 2: Posterior Estimates for the Pause Before Models. The factor "Intro Verb" captures the effect that the presence of a speech verb has on the occurrence of a pause in the same position. Other factors are explained in the main text.

presence of an ending verb has a strong negative effect (mean ≈ -4.28), meaning that the probability of a pause is reduced to just 14 % when an ending verb is present. The effects related to the speaker's characteristics, age and sex, are not conclusive in this context. Taken together, these results indicate that the presence of a speech verb and the presence of a pause are inversely related, both at the beginning and even more strongly at the end of speech reports.

Figure 3: Posterior Estimates for the Pause After Models. The factor "Ending Verb" captures the effect that the presence of a speech verb has on the occurrence of a pause in the same position. Other factors are explained in the main text.

Figures 2 and 3 also include the results that address the effect of basic word order on pause probabilities (Hypothesis 2.2). Regarding pauses before speech reports (Figure 2), we find strong evidence that having no dominant word order has a positive effect (mean ≈ 1.87), indicating that in Arapaho and Mojeno there is a higher probability of pauses before speech reports than in OV and VO languages. On the other hand, there is no evidence for an effect of VO word order on the probability of pauses before speech reports, nor are there credible effects of interactions between

the presence of an introductory verb and any word order. For pauses after speech reports (Figure 3), we found no evidence for more pausing in languages with no dominant word order, and also not for VO languages. However, we did find evidence for interactions between word order and ending verbs: For languages with no dominant word order, there is strong evidence for a positive effect (mean ≈ 1.56) and for VO languages there is weak evidence for a likewise positive effect (mean ≈ 1.29). This indicates that, depending on the basic word order of a language, the impact of the presence of an ending verb on the probability of a silent pause after the speech report varies: VO languages and languages with no dominant word order tend to have more pauses in this context, compared to OV languages. However, the result regarding VO languages should be interpreted with caution since Mojeño is the only VO language that uses any ending verbs in our corpus. Overall, our results thus indicate, firstly, that languages with no dominant word order clearly tend to pause more both before and after speech reports. Secondly, there is only weak evidence, and only for VO languages, for conventionalization of pausing in contexts where speech verbs do not occur according to the basic word order, i.e. before in OV languages and after in OV languages, independently of complementarity.

Next, we present two more results that further contextualize our main results. Firstly, Figure 4 provides language-specific results of posterior predictions from the Bayesian models, comparing pause probabilities before, in blue, and after, in red, depending on the presence versus absence of an introductory verb or an ending verb. These results indicate that complementarity between pauses and speech verbs prevails across individual languages and contexts (before vs. after). The results appear to also indicate potential word order effects. In OV languages there appears to be a relatively high probability of a pause preceding speech reports, i.e. in the position where typically there is no speech verb, even if a speech verb happens to be present (dark blue bars for OV languages). In VO languages, on the other hand, there appears to be typically a relatively high probability of a pause following speech reports, i.e. in the position where typically there is no speech verb, even if a speech verb happens to be present (dark red bars for VO languages). Figure 4 also illustrates variation across languages and word order types, emphasizing that while the overall patterns support complementarity and the influence of basic word order, individual languages differ in the magnitude of these effects. Note, however, the uncertainty of some of the individual results, reflected in excessively wide error bars that represent the 95 % HDI credible intervals associated with the model, e.g. for pause probabilities in the presence of introductory verbs in Ruuli and N‖ng. This is partially due to the fact that the models infer the probabilities of a pauses for every language in every position, even if they do not have ending verbs in the corpus, as is the case for Ruuli and N‖ng, for which there are no ending verbs attested in our corpora.
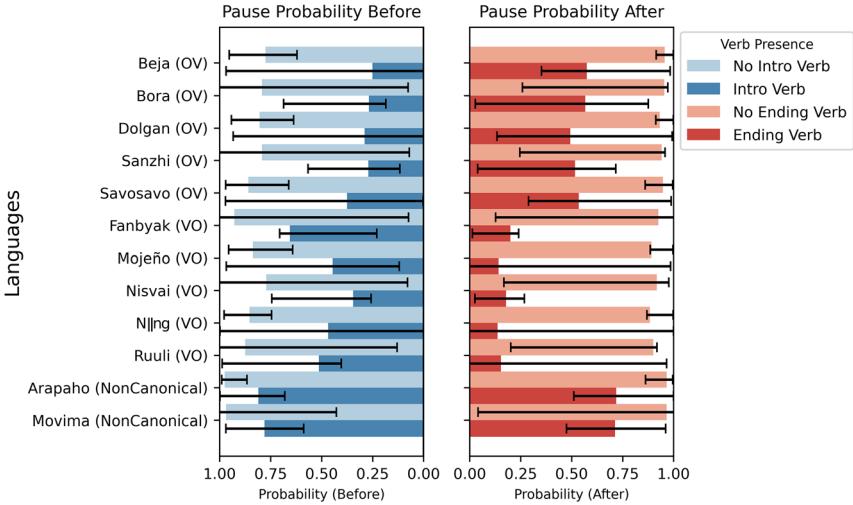
Figure 4: Pause probabilities in the presence versus absence of speech verbs before and after speech reports in 12 individual languages.

Finally, we report descriptive results on pause durations in Figure 5. These show that before speech reports, the durations of pauses tend to be longer when there is no introductory verb (mean ≈ 0.97s) compared to when there is an introductory verb (mean ≈ 0.81s). In positions after speech reports, when an ending verb is absent, the mean pause duration is also relatively long (mean ≈ 0.78s), but when an ending verb is present, pauses are much shorter (mean ≈ 0.38s). A Bayesian model assessing the standardized pause duration also indicates that the presence of an introductory verb has a credible negative effect on introductory pause duration (mean = -0.220, HDI close to zero), albeit marginal. For ending verbs, the Bayesian analysis is clearer: a negative slope of about -0.569 indicates that ending verbs are associated with shorter pauses at the offset. Taken together, these results provide additional evidence for the complementarity hypothesis, and they also again show that pauses before and after speech reports differ. On the one hand, pauses before speech reports tend to be longer, regardless of the presence or absence of a speech verb, compared to pauses after speech reports. On the other hand, in the position between speech reports and ending verbs it is not only relatively rare for a pause to occur, but if pauses do occur there, they are also remarkably short.
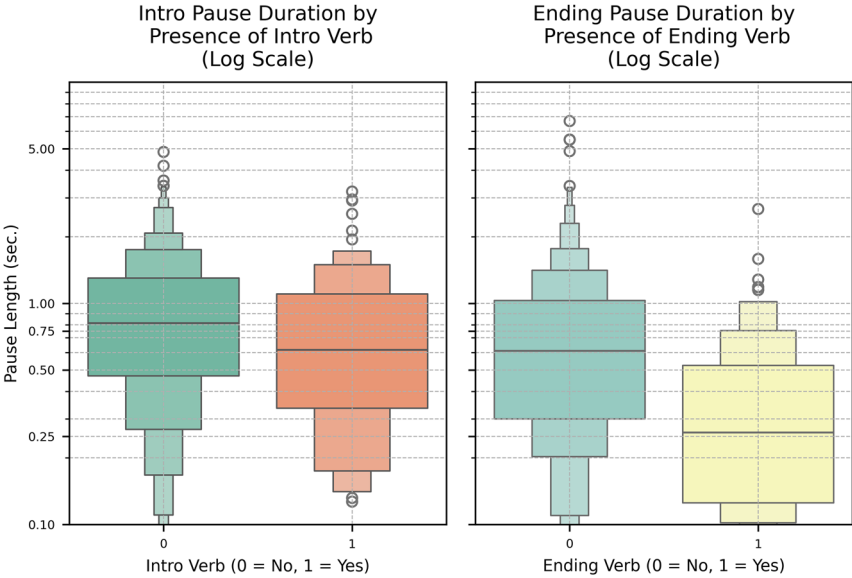
Figure 5: Duration of silent pauses in the absence versus presence of introductory and ending verbs.

4.4 Pauses before and after middle verbs

As noted in Section 4.1, OV languages and languages with no dominant order employ – in some cases many – middle verbs, in addition to ending and introductory verbs. Figures 2 and 3 suggest that the presence of middle verbs does not have an effect on pauses before or after the speech report as whole, as expected. In the current section, we briefly report on pauses before and after middle verbs. Since there are only relatively few middle verbs in most languages, except in Bora and Dolgan, we do not extend the modeling approach to the study of pauses before and after them, but we simply provide basic descriptive statistics (Table 5). These very preliminary results indicate that in most languages, middle verbs are often followed by pauses, i.e. a pause occurs between the middle verb and the following speech report. Only rarely (except in Arapaho) are middle verbs preceded by pauses, i.e. pauses only rarely intervene between the end of a speech report and a middle verb.

These preliminary results also indicate that middle verbs behave like ending verbs in terms of pausing, in that they follow preceding speech reports with little pausing, as can be seen from comparison of the column "pause before verb" in Table 5 with the estimated probability of, on average, 14 % for pauses before ending

Table 5: Percentage of pauses before and after middle verbs, indicating differences of more than 10 % by "<", and more than 25 % by "<<".

Language	BWO	Pause before verb		Pause after verb
Arapaho	No dom.	62.5 % (5/8)	<<	87.5 % (7/8)
Movima	No dom.	14.3 % (1/7)	<<	71.4 % (5/7)
Beja	S**OV**	0.0 % (0/13)	<	23.1 % (3/13)
Bora	S**OV**	8.5 % (6/71)	<<	50.7 % (36/71)
Dolgan	S**OV**	4.5 % (2/44)	<<	52.3 % (23/44)
Sanzhi	S**OV**	7.1 % (1/14)	<<	35.7 % (5/14)
Savosavo	S**OV**	n/a (1/1)		n/a (1/1)
Fanbyak	S**VO**	n/a (0/0)		n/a (0/0)
Mojeño	S**VO**	n/a (1/3)		n/a (3/3)
Nisvai	S**VO**	n/a (0/0)		n/a (0/0)
Nǁng	S**VO**	n/a (0/0)		n/a (0/0)
Ruuli	S**VO**	n/a (0/0)		n/a (0/0)

verbs (Figure 3). They also behave like introductory verbs in that pauses are more likely to intervene between the middle verb and a following speech report, as can be seen from comparison of the column "pause before verb" in Table 5 with the esti-mated probability of, on average, 48 % for pauses after introductory verbs (Figure 2). The two languages with no dominant word order appear to be characterized by a lot of pausing in the context of middle verbs, too, although these results are particularly uncertain because both languages use particularly few middle verbs. Additionally, we observed that in terms of their structure, as discussed in Section 4.1, middle verbs also tend to resemble ending verbs in that they tend to head short, one-word phrases consisting only of the speech verb (see examples (4)–(5), above), without further elaboration regarding participants or circumstances, etc., which is typical of introductory verbs.

5 Discussion

Our study has revealed high degrees of variability in the construal and prosodic phrasing of speech reports, both within and across languages. The placement of speech verbs and, in particular, pause occurrences rarely displays truly categorical behavior. However, both regarding position and pauses, the overall patterns are far from random, and they are closely connected with the basic word order of a language.

Our findings on the order of speech verbs and speech reports support our first hypothesis, according to which speech reports tend to follow speech verbs in VO languages and mostly precede them in OV languages: In our sample, four out of five VO languages used exclusively introductory verbs, while the fifth one displayed an entirely different behavior. Among our five OV languages, all of them used primarily ending verbs, but also introductory (and middle) verbs to varying degrees. These findings closely match Güldemann's (2008: 193–194, 210–219) findings from corpus counts on a sample of 39 African languages. Interestingly, he also found exceptions in each category, notably two out of 20 VO languages, Lamang and Nguni, use slightly more postposed than preposed speech verbs, similar to the exceptional VO language Mojeño from our sample. This indicates that there may be additional, so far unknown, factors involved in the sequential ordering of speech reports. One possibility is that these languages are undergoing or have recently undergone shifts in basic word order, possibly triggered by areal pressure.

The ordering patterns we found are also strikingly similar to the relationship between the order of verbs and nominal objects and that of complement-taking verbs and complement clauses reported by Schmidtke-Bode and Diessel (2017: 10–12) for a stratified sample of 100 languages, even though their figures are not directly comparable to ours because they analyzed grammatical ordering rules for complement clauses, unlike the corpus counts we provided. They found that in VO languages, complements almost always follow complement taking verbs. Among the OV languages they studied, 51.7 % have exclusively preverbal complements, 23.3 % have postverbal complements only and 25 % allow postverbal complements. Counts on individual constructions, rather than languages as a whole, confirm this pattern. Note that all of their counts include "utterance verbs", although they do not provide separate figures for these and they do not differentiate between direct and indirect speech.

As noted in the introduction, the relation between VO versus OV and the ordering of complement clauses and speech reports is also reminiscent of the order of relative clauses and their head nouns (Dryer 2013). That is, relative clauses almost always follow nouns in VO languages but can precede or follow them in OV languages. The typology of the sequential ordering of these three constructions thus seems to result from the same set of competing motivations (Hawkins 2004: 205ff.). A pressure for consistency in the order of heads and dependents, on the one hand, and a pressure for "heavy last", on the other hand, which results in mixed ordering patterns for OV languages.

In addition to introductory and ending verbs, the current study also focused on middle verbs. Here, we found many middle verbs in OV languages. We interpret their use as a strategy in these languages to insert a speech verb early, despite pressure to place it at the end of a speech report. Regarding middle verbs, our results clearly

differ from those of Güldemann's (2008) study on African languages: He found only a few instances of middle verbs ("intraposed" in his terminology), and these were equally rare in VO and OV languages. But Güldemann (2008) found other strategies to mitigate the inconvenience of using ending verbs only, which violates the heavy-last principle. These including "an anticipatory quote proform before the matrix [speech] verb, apparently to saturate its referential and possibly syntactic valency" (Güldemann 2008: 214) and various types of bipartite "circumposed" speech verbs. Neither strategy was found in the languages considered in the current study. It is unclear so far whether the African languages studied by Güldemann (2008) are really different from our world-wide sample regarding such middle and circumposed verbs, as well as "anticipatory quote proforms", or whether the difference is due to differences in the methods used.

The two languages with no dominant word order in our sample, Movima and Arapaho, behave like OV languages in terms of the relatively variable use of each ordering type, i.e. introductory, middle, and ending verbs, contrasting with the more homogenous pattern displayed by most VO languages, which exclusively use introductory verbs, even though among these languages, too, exceptions are possible, as the case of Mojeño in our sample showed.

Regarding the occurrence of pauses, we found, as expected, even more variability within languages than regarding the position of speech verbs. Our findings revealed evidence for various factors involved in predicting the occurrence of pauses before and after speech reports. Firstly, we found strong evidence for our Hypothesis 2.1, according to which speech verbs and pauses display complementarity, i.e., silent pauses will occur at those edges of speech reports where speech verbs are absent. Our results indicate that before speech reports, the probability of a pause to occur is 88 % in the absence of an introductory verb, but only 48 % when an introductory verb is present. After speech reports, there is a 92 % probability for a pause to occur in the absence of a speech verb, but just 14 % when an ending verb is present. The fact that complementarity is much stronger for ending verbs than for introductory verbs is a first indication of the important differences between these two, that will be discussed further below.

Secondly, we found some evidence for Hypothesis 2.2, according to which silent pauses will occur more often before than after speech reports in OV languages and more often after than before speech reports in VO languages, also independently of the presence of a speech verb. If true, this would suggest that the basic word order of a language influences the probability of pauses beyond mere complementarity. However, the evidence for higher pause probabilities before speech reports in OV and after them in VO languages was weak and not fully reliable. We thus hesitate to conclude that prosodic phrasing of speech reports through pauses would be to some extent conventionalized in a language, beyond the mere need for boundary marking

in the absence of speech verbs. Future research will have to investigate whether the lack of stronger evidence for such direct word order effects reflects a true lack of relationship or is due to scarcity of cross-linguistic coverage in our data.

Our findings further show that pauses' occurrence patterns before and after speech reports are not the mirror image of each other and relate differently to the presence versus absence of speech verbs as well as word order. Firstly, there are many pauses preceding speech reports even in the presence of speech verbs (see also Güldemann 2008: 222). One reason for this may be that marking the beginning of speech reports may be more important than marking its end, to signal the special nature of speech reports early on in discourse. Another might be that these pauses also reflect the effort that speakers have to make to prepare their performance of direct speech. And secondly, we noted that preposed speech verbs tend to head structurally more complex phrases, which may necessitate intonation breaks and pauses between these phrases and the following speech report. We noted that this higher complexity most often involves explicit mentions of discourse participants. As such, the asymmetry between introductory and ending verbs is also consistent with sequential discourse constraint of introducing actors early on in a narrative (also called exposition or orientation, see for instance Adam 2011; Labov 2010). While we were not able to directly investigate the effect of the length of the verb phrase containing the introductory verb on pausing, we were able to show increasing probabilities for pausing before speech reports the longer the stretches of speech since the previous pause are. As a final observation regarding the asymmetry in pausing before and after speech verbs, from a processing perspective (Himmelmann 2014), it makes more sense to pause before a relatively long, complex and probably largely unpredictable constituent, like a speech report, than before a relatively short, possibly minimally complex and in any case highly predictable constituent like a speech verb.

How do pause probabilities at the edges of speech reports relate to the pause probabilities at different kinds of clause boundaries? Peck and Becker (2024) reported between 29 % and 57 % pauses at dependent clause boundaries and between 50 % and 75 % at main clause boundaries. We may compare speech report boundaries with no speech verbs with main clause boundaries – assuming that a speech report is not syntactically linked to any element in the surrounding discourse other than a speech verb. Here, our results indicated a mean probability of pauses across languages of about 88 % before speech reports and of 92 % after speech reports. Boundary marking by pauses for speech reports is thus even more likely, especially after speech reports, than for main clause boundaries in general. This may be interpreted as an effect of the extra need for "reorientation" of direct speech. On the other hand, we can compare boundaries of speech reports in the presence of speech verbs with dependent clause boundaries, in as much as speech reports can

be compared to complement clauses. For these contexts results indicate pause probabilities of 48.% before speech reports and 14 % after speech reports. These figures for introductory verbs fall within the range reported by Peck and Becker (2024) for dependent clause boundaries, but ending verbs display a special behavior also in the context of pausing at dependent clause boundaries.

Regarding middle verbs, we noted that these are similar to ending verbs in terms of structure and pausing, including that, in our data, they only occur in languages that also use ending verbs, so for our overall conclusions, we may treat them together: Ending verbs – a feature almost exclusive to OV languages – are linguistic structures that are special in various respects. Firstly, regarding their co-occurrence with other types of speech verbs, we may formulate the following universal:

Universal: If a language has ending verbs, it also has introductory verbs, but not vice versa.

Secondly, regarding their structural and prosodic features, they tend to be short, phonetically reduced forms, as already noted by Güldemann (2008: 197–198). Our preliminary results suggest that in addition, they regularly prosodically phrase with the preceding speech report, as indicated by only rare occurrence of pauses, and – if a pause is present – the shortness of pauses. This adds credibility to hypotheses about the grammaticalization of postposed speech verbs into quotative suffixes discussed earlier in the literature (Güldemann 2008: 224).

As a final observation we note that from our data potentially emerges a third type of language, with no dominant word order, many speech reports without any speech verb, and many pauses both preceding and following speech reports, even in the presence of speech verbs, although for the time being this proposal is highly speculative as it is based on only two languages.

6 Conclusions

On the one hand, Klewitz and Couper-Kuhlen (1999: 482–483) state that "the prosodic marking of reported speech in spoken discourse […] is a stylistic device rather than a norm: It may be used to signal reported speech or not, depending on speakers' local goals and strategic choices". On the other hand, de Brabanter (2023) argues that prosodic marking is actually the norm because speech reports have to be systematically distinguished from the surrounding discourse. The results on speech reports in natural discourse in 12 languages from around the world presented here show that pausing at the edges of speech reports is fairly variable in most contexts, but overall far from random, and dependent on the basic word order of a language in

non-trivial ways. Reaching these conclusions was made possible by applying a comparative-corpus approach to time-aligned corpus data from a typologically diverse set of languages that has only recently become available. As a particularly relevant line of future research emerging from our study, we suggest the systematic inclusion of interjections, as these elements closely interact with both speech verb type – especially the use of middle verbs after interjections – and pausing patterns. Further questions for future research include how other prosodic markers of reported speech, such as pitch movements, could complement the result reported on in the present study, and whether and how results obtained from conventionalized narratives can generalize to other speech styles, in particular conversational speech.

Author contributions: Both authors contributed equally to the study. FS and JA designed the study, JA coded most of the data and developed the processing, JA and FS analysed the data, and FS and JA wrote the paper.

Data availability: The data supporting the results can be found at https://doi.org/10.5281/zenodo.10870944 (Aznar 2024).

References

Abril-Pla, Oriol, Virgile Andreani, Colin Carroll, Larry Dong, Christopher J. Fonnesbeck, Maxim Kochurov, Ravin Kumar, et al. 2023. PyMC: A modern, and comprehensive probabilistic programming framework in python. *PeerJ Computer Science* 9. e1516.

Adam, Jean-Michel. 2011. *Les textes, types et prototypes: séquences descriptives, narratives, argumentatives, explicatives, dialogales et genres de l'injonction-instruction* [Texts, types and prototypes: descriptive, narrative, argumentative, explanatory, dialogical sequences and injunction-instruction genres]. Paris: A. Colin.

Aikhenvald, Alexandra Y. 2008. Semi-direct speech: Manambu and beyond. *Language Sciences* 30(4). 383–422.

Aznar, Jocelyn. 2019. *Narrer une nabol : La production des textes nisvais en fonction de l'âge et de la situation d'énonciation, Malekula, Vanuatu* [Telling a nabol: the production of Nisvai texts according to the age and the enunciative situation, Malekula, Vanuatu]. Marseille: EHESS.

Aznar, Jocelyn. 2020. D'un corpus à l'autre: D'une étude reproductible et portable du discours direct nisvai à la comparaison linguistique [From one corpus to another: from a reproducible and portable study of direct speech in Nisvai to the linguistic comparison]. In Thierry Poibeau, Yannick Parmentier &

Emmanuel Schang (eds.), *LIFT 2020. Actes des 2èmes journées scientifiques du Groupement de Recherche Linguistique Informatique Formelle et de Terrain (LIFT), 10-11 décembre 2020 Montrouge, France (Virtuel)*, 42–52. Available at: https://lift2020.sciencesconf.org.

Aznar, Jocelyn. 2022. Nisvai DoReCo dataset. In Frank Seifart, Ludger Paschen & Matthew Stave (eds.), *Language documentation reference corpus (DoReCo) 1.0*. Berlin & Lyon: Leibniz-Zentrum Allgemeine Sprachwissenschaft & laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2).

Aznar, Jocelyn. 2024. Notebook for the crosslinguistic study of direct speech, word order and silent pause. Zenodo. https://doi.org/10.5281/zenodo.10870944.

Aznar, Jocelyn & Núria Gala. 2020. The Nisvai corpus of oral narrative practices from Malekula (Vanuatu) and its associated language resources. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, 2649–2656. Marseille, France: European Language Resources Association. Available at: https://aclanthology.org/2020.lrec-1.323.

Aznar, Jocelyn & Frank Seifart. 2020. RefCo: An initiative to develop a set of quality criteria for fieldwork corpora. In Thierry Poibeau, Yannick Parmentier & Emmanuel Schang (eds.), *LIFT 2020. Actes des 2èmes journées scientifiques du Groupement de Recherche Linguistique Informatique Formelle et de Terrain (LIFT), 10–11 décembre 2020 Montrouge, France (Virtuel)*, 94–100. Available at: https://lift2020.sciencesconf.org.

Aznar, Jocelyn & Frank Seifart. 2022. *The RefCo Toolkit*. Zenodo. https://doi.org/10.5281/zenodo.7380448.

Barth, Danielle, Nicholas Evans, I. Wayan Arka, Henrik Bergqvist, Diana Forker, Sonja Gipper, Gabrielle Hodge, et al. 2021. Language vs individuals in cross-linguistic corpus typology. In Geoffrey Haig, Stefan Schnell & Frank Seifart (eds.), 179–232. Honolulu: University of Hawai'i Press.

Behagel, Otto. 1909. Beziehungen zwischen Umfang und Reihenfolge von Satzgliedern [The relationship between the length and order of constituents]. *Indogermanische Forschungen* 25. 110–142.

Blyth, Carl, Sigrid Recktenwald & Jenny Wang. 1990. I'm like, "Say What?!": A new quotative in American oral narrative. *American Speech* 65(3). 215.

Buchstaller, Isabelle & Ingrid van Alphen (eds.). 2012. *Quotatives: Cross-linguistic and cross-disciplinary perspectives*. Converging Evidence in Language and Communication Research, v. 15. Amsterdam & Philadelphia: John Benjamins.

Cervone, Alessandra, Catherine Lai, Silvia Pareti & Peter Bell. 2015. Towards automatic detection of reported speech in dialogue using prosodic cues. *Proceedings of Interspeech* 2015. 3061–3065.

Coulmas, Florian. 1986. Introduction. In Florian Coulmas (ed.), *Direct and indirect speech*, 1–28. Berlin: Mouton de Gruyter.

Couper-Kuhlen, Elizabeth. 1999. Coherent voicing: On prosody in conversational reported speech. In Wolfram Bublitz, Uta Lenk & Eija Ventola (eds.), *Coherence in spoken and written discourse. How to create it and how to describe it* (Pragmatics & Beyond New Series 63), 11–34. Amsterdam: John Benjamins.

Cowell, Andrew. 2022. Arapaho DoReCo dataset. In Frank Seifart, Ludger Paschen & Matthew Stave (eds.), *Language documentation reference corpus (DoReCo) 1.0*. Berlin & Lyon: Leibniz-Zentrum Allgemeine Sprachwissenschaft & laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2).

Cowell, Andrew & Alonzo Moss. 2008. *The Arapaho language*. Boulder: University Press of Colorado.

Culy, Christopher. 1997. Logophoric pronouns and point of view. *Linguistics* 35(5). 845–860.

Däbritz, Chris Lasse, Nina Kudryakova, Eugénie Stapert & Alexandre Arkhipov. 2022. Dolgan DoReCo dataset. In Frank Seifart, Ludger Paschen & Matthew Stave (eds.), *Language documentation reference corpus (DoReCo) 1.0*. Berlin & Lyon: Leibniz-Zentrum Allgemeine Sprachwissenschaft & laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2).

De Brabanter, Phillippe. 2023. Quotation does not need marks of quotation. *Linguistics* 61(2). 285–316.

Demers, Monique. 2009. Prosodie et discours. Le cas du discours rapporté en français québécois oral spontané [Prosody and discourse. The case of reported speech in Quebec French spontaneous speaking]. *Revue Québécoise de Linguistique* 26(1). 27–50.

Dik, Simon C. 1997. *The theory of functional grammar, Part 1: The structure of the clause*. Berlin, New York: Mouton de Gruyter.

Donnelly, Seamus, Caroline Rowland, Franklin Chang & Evan Kidd. 2024. A comprehensive examination of prediction-based error as a mechanism for syntactic development: Evidence from syntactic priming. *Cognitive Science* 48(4). e13431.

Dryer, Matthew S. 1980. The positional tendencies of sentential noun phrases in universal grammar. *Canadian Journal of Linguistics/Revue Canadienne de Linguistique* 25(2). 123–196.

Dryer, Matthew S. 1989. Large linguistic areas and language sampling. *Studies in Language* 13(2). 257–292.

Dryer, Matthew S. 2013. Order of subject, object and verb. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available at: https://wals.info/chapter/81.

Dryer, Matthew S. & Martin Haspelmath (eds.). 2013. *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available at: http://wals.info/.

ELAN Developers. 2021. *ELAN (Version 6.1) [Computer software]*. Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. Available at: https://archive.mpi.nl/tla/elan.

Engelmann, Felix, Sonia Granlund, Joanna Kolak, Marta Szreder, Ben Ambridge, Julian Pine, Anna Theakston & Elena Lieven. 2019. How the input shapes the acquisition of verb morphology: Elicited production and computational modelling in two highly inflected languages. *Cognitive Psychology* 110. 30–69.

Evans, Nicholas. 2013. Some problems in the typology of quotation: A canonical approach. In Dunstan Brown, Marina Chumakina & Greville G. Corbett (eds.), *Canonical morphology and syntax*, 66–98. Oxford University Press.

Forker, Diana. 2019. Sanzhi Dargwa dictionary. *Dictionaria* 5. 1–5533.

Forker, Diana & Nils Norman Schiborr. 2022. Sanzhi Dargwa DoReCo dataset. In Frank Seifart, Ludger Paschen & Matthew Stave (eds.), *Language documentation reference corpus (DoReCo) 1.0*. Berlin & Lyon: Leibniz-Zentrum Allgemeine Sprachwissenschaft & laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2).

Franjieh, Michael. 2022. Fanbyak DoReCo dataset. In Frank Seifart, Ludger Paschen & Matthew Stave (eds.), *Language documentation reference corpus (DoReCo) 1.0*. Berlin & Lyon: Leibniz-Zentrum Allgemeine Sprachwissenschaft & laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2).

Galucio, Ana. 2024. Quote constructions in the Sakurabiat language. *Language Documentation and Description* 23(2). 1–16.

Genetti, Carol. 2011. Direct speech reports and the cline of prosodic integration in Dolakha Newar. *Himalayan Linguistics* 10(1). 55–76.

Grosjean, François & Maryann Collins. 1979. Breathing, pausing and reading. *Phonetica* 36(2). 98–114.

Güldemann, Tom. 2008. *Quotative indexes in African languages. A synchronic and diachronic survey*. Empirical Approaches to Language Typology [EALT]. Berlin, New York: Mouton de Gruyter.

Güldemann, Tom, Martina Ernszt, Sven Siegmund & Alena Witzlack-Makarevich. 2022. Nǁng DoReCo dataset. In Frank Seifart, Ludger Paschen & Matthew Stave (eds.), *Language documentation reference corpus (DoReCo) 1.0*. Berlin & Lyon: Leibniz-Zentrum Allgemeine Sprachwissenschaft & laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2).

Günthner, Susanne. 1999. Polyphony and the 'layering of voices' in reported dialogues: An analysis of the use of prosodic devices in everyday reported speech. *Journal of Pragmatics* 31(5). 685–708.

Hagège, Claude. 1974. Les pronoms logophoriques (Exemples en mundang, tuburi, éwé et langues oubanguiennes| cas du japonais et du coréen) [Logophoric pronouns (with examples from Mundang, Tubiru, Éwé and Ubangi languages: the case of Japanese and Korean)].|. *Bulletin de la Société de Linguistique de Paris* 69(1). 287–310.

Hammarström, Harald, Robert Forkel, Martin Haspelmath & Sebastian Bank. 2022. *Glottolog 4.7*. Leipzig. https://doi.org/10.5281/zenodo.7398962.

Hanote, Sylvie. 2015. Discours direct : syntaxe et prosodie [Direct speech: syntax and prosody]. *E-rea*, no 12.2. https://doi.org/10.4000/erea.4244.

Haude, Katharina. 2006. *A grammar of Movima*. Nijmegen University Ph.D. dissertation. Available at: https://hdl.handle.net/2066/41395.

Haude, Katharina. 2022. Movima DoReCo dataset. In Frank Seifart, Ludger Paschen & Matthew Stave (eds.), *Language documentation reference corpus (DoReCo) 1.0*. Berlin & Lyon: Leibniz-Zentrum Allgemeine Sprachwissenschaft & laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2).

Hawkins, John A. 1994. *A performance theory of order and constituency*. Cambridge: Cambridge University Press.

Hawkins, John A. 2004. *Efficiency and complexity in grammars*. Oxford: Oxford University Press.

Henderson, Alan, Frieda Goldman-Eisler & Andrew Skarbek. 1965. Temporal patterns of cognitive activity and breath control in speech. *Language and Speech* 8(4). 236–242.

Himmelmann, Nikolaus P. 2014. Asymmetries in the prosodic phrasing of function words: Another look at the suffixing preference. *Language* 90(4). 927–960.

Himmelmann, Nikolaus P. & D. Robert Ladd. 2008. Prosodic description: An introduction for fieldworkers. *Language Documentation & Conservation* 2(2). 244–274.

Izre'el, Shlomo & Amina Mettouchi. 2015. Transcriptional strategies and prosodic units: Representation of speech in CorpAfroAs. In Amina Mettouchi, Martine Vanhove & Dominique Caubet (eds.), *Corpus-based studies of lesser-described languages: The CorpAfroAs corpus of spoken AfroAsiatic languages* (Studies in Corpus Linguistics 68), 13–41. Amsterdam: John Benjamins Publishing Company.

Jacewicz, Ewa, Robert A. Fox, Caitlin O' Neill & Joseph Salmons. 2009. Articulation rate across dialect, age, and gender. *Language Variation and Change* 21(2). 233–256.

Jansen, Wouter, Michelle L. Gregory & Jason M. Brenier. 2001. Prosodic correlates of directly reported speech: Evidence from conversational speech. In *Proceedings of the ISCA Workshop on Prosody in Speech Recognition and Understanding*, paper 14. Available at: https://www.isca-archive.org/prosody_2001/jansen01_prosody.html.

Jordanoska, Izabela, Ekaterina Aplonova, Guillaume Guitang, Olga Kuznetsova & Songfolo Lacina Silue. 2022. Saying without 'say': A corpus-based approach to the typology of non-canonical speech reports. Paper given at the Workshop "Spoken- and Signed-language Corpus Studies in Linguistic Typology", 14th International Conference of the Association for Linguistic Typology (ALT 2022). USA: University of Texas at Austin.

Klewitz, Gabriele & Elizabeth Couper-Kuhlen. 1999. Quote – unquote? The role of prosody in the contextualization of reported speech sequences. *Pragmatics* 9(4). 459–485.

Kluyver, Thomas, Benjamin Ragan-Kelley, Fernando Pérez, Brian Granger, Matthias Bussonnier, Jonathan Frederic, Kyle Kelley, et al. 2016. Jupyter Notebooks – a publishing format for reproducible computational workflows. In Fernando Loizides & Birgit Schmidt (eds.), *Positioning and Power in Academic Publishing: Players, Agents and Agendas. Proceedings of the 20th International Conference on Electronic Publishing*, 87–90. Amsterdam, Berlin, Washington DC: IOS Press.

Kumarage, Shanthi, Seamus Donnelly & Evan Kidd. 2022. Implicit learning of structure across time: A longitudinal investigation of syntactic priming in young English-acquiring children. *Journal of Memory and Language* 127. 104374.

Labov, William. 2010. Oral narratives of personal experience. In Patrick Colm Hogan (ed.), *The Cambridge encyclopedia of the language sciences*, 586–588. Cambridge: Cambridge University Press.

Leandri, Sylvie. 1993. Prosodic aspects of reported speech. Working papers/Lund University, Department of Linguistics and Phonetics 41. Available at: https://journals.lub.lu.se/LWPL/article/view/17180.

Levshina, Natalia, Savithry Namboodiripad, Marc Allassonnière-Tang, Mathew Kramer, Luigi Talamo, Annemarie Verkerk, Sasha Wilmoth, et al. 2023. Why we need a gradient approach to word order. *Linguistics* 61(4). 825–883.

Longacre, Robert E. 2007. Sentences as combinations of clauses. In Timothy Shopen (ed.), *Language typology and syntactic description: Volume 2: Complex constructions*, 2nd edn., 372–420. Cambridge: Cambridge University Press.

Malibert, Il-Il & Martine Vanhove. 2015. Quotative constructions and prosody in some Afroasiatic languages: Towards a typology. In Amina Mettouchi, Martine Vanhove & Dominique Caubet (eds.), *Corpus-based Studies of lesser-described languages: The CorpAfroAs corpus of spoken AfroAsiatic languages*. (Studies in Corpus Linguistics 68), 117–169. Amsterdam: John Benjamins.

McGregor, William. 1994. The grammar of reported speech and thought in Gooniyandi. *Australian Journal of Linguistics* 14(1). 63–92.

Michno, Jeff. 2021. The functions of prosody in the reported speech of rural Southwestern Nicaraguans. *Semas* 2(3). 45–67.

Munro, Pamela. 1982. On the transitivity of 'say' verbs. In Paul J. Hopper & Sandra A. Thompson (eds.), *Studies in transitivity*. (Syntax and Semantics 15), 301–318. New York: Academic Press.

Nikitina, Tatiana & Anna Bugaeva. 2021. Logophoric speech is not indirect: Towards a syntactic approach to reported speech constructions. *Linguistics* 59(3). 609–633.

Norrick, Neal R. 2014. Interjections. In Christoph Rühlemann & Karin Aijmer (eds.), *Corpus pragmatics: A handbook*, 249–276. Cambridge: Cambridge University Press.

Paschen, Ludger, François Delafontaine, Christoph Draxler, Susanne Fuchs, Matthew Stave & Frank Seifart. 2020. Building a time-aligned cross-linguistic reference corpus from language documentation data (DoReCo). In *Proceedings of the 12th Language Resources and Evaluation Conference*, 2657–2666. Marseille, France: European Language Resources Association. Available at: https://www.aclweb.org/anthology/2020.lrec-1.324.

Pearson, Hazel. 2015. The interpretation of the logophoric pronoun in Ewe. *Natural Language Semantics* 23(2). 77–118.

Peck, Naomi & Laura Becker. 2024. Syntactic pausing? Re-examining the associations. *Linguistics Vanguard* 10(1). 223–237.

Peck, Naomi, Kirsten Culhane & Maria Vollmer. 2021. Comparing cues: A mixed methods study of intonation unit boundaries in three typologically diverse languages. Paper presented at the AG 10a Prosodic boundary phenomena at the 43rd Annual Conference of the German Linguistic Society, 23-26 February 2021, Freiburg.

Rose, Françoise. 2014. Mojeño Trinitario. In Pieter Muysken & Mily Crevels (eds.), *Oriente (Lenguas de Bolivia 3)*, 59–97. La Paz: Plural Editores.

Rose, Françoise. 2022. Mojeño Trinitario DoReCo dataset. In Frank Seifart, Ludger Paschen & Matthew Stave (eds.), *Language documentation reference corpus (DoReCo) 1.0*. Berlin & Lyon: Leibniz-Zentrum Allgemeine Sprachwissenschaft & laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2).

Schmidtke-Bode, Karsten & Holger Diessel. 2017. Cross-linguistic patterns in the structure, function, and position of (object) complement clauses. *Linguistics* 55(1). 1–38.

Schnell, Stefan & Nils Norman Schiborr. 2022. Crosslinguistic corpus studies in linguistic typology. *Annual Review of Linguistics* 8. 171–191.

Seifart, Frank. 2022. Bora DoReCo dataset. In Frank Seifart, Ludger Paschen & Matthew Stave (eds.), *Language documentation reference corpus (DoReCo) 1.0*. Berlin & Lyon: Leibniz-Zentrum Allgemeine Sprachwissenschaft & laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2).

Seifart, Frank, Ludger Paschen & Matthew Stave (eds.). 2022. *Language documentation reference corpus (DoReCo) 1.0*. Berlin & Lyon: Leibniz-Zentrum Allgemeine Sprachwissenschaft & laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2).

Sørensen, Marie-Louise Lind & Alena Witzlack-Makarevich. 2020. Clausal complementation in Ruuli (Bantu, JE103). *Studies in African Linguistics* 49(1). 84–110.

Spronck, Stef. 2012. Minds divided: Speaker attitudes in quotatives. In Isabelle Buchstaller & Ingrid van Alphen (eds.), *Converging evidence in language and communication research*, vol. 15, 71–116. Amsterdam: John Benjamins Publishing Company.

Spronck, Stef & Tatiana Nikitina. 2019. Reported speech forms a dedicated syntactic domain. *Linguistic Typology* 23(1). 119–159.

Stapert, Eugénie. 2013. *Contact-induced change in Dolgan: An investigation into the role of linguistic data for the reconstruction of a people's (pre)history*. Ph.D. dissertation, Leiden University. Available at: https://hdl.handle.net/1887/21798.

Tannen, Deborah. 1986. Introducing constructed dialogue in Greek and American conversational and literary narrative. In Florian Coulmas (ed.), *Direct and indirect speech*, 311–332. Berlin: Mouton de Gruyter.

Torreira, Francisco, Sara Bögels & Stephen C. Levinson. 2015. Breathing for answering: The time course of response planning in conversation. *Frontiers in Psychology* 6. 1–11.

Vanhove, Martine. 2017. *Le bedja* (Les Langues du Monde 68). Leuven, Paris: Peeters.

Vanhove, Martine. 2022. Beja DoReCo dataset. In Frank Seifart, Ludger Paschen & Matthew Stave (eds.), *Language documentation reference corpus (DoReCo) 1.0*. Berlin & Lyon: Leibniz-Zentrum Allgemeine Sprachwissenschaft & laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2).

Voloshinov, Valentin Nikolaevich. 1986. *Marxism and the philosophy of language*. Cambridge, Mass: Harvard University Press.

Wegener, Claudia. 2012. *A grammar of Savosavo*. (Mouton Grammar Library 61). Berlin: De Gruyter Mouton.

Wegener, Claudia. 2022. Savosavo DoReCo dataset. In Frank Seifart, Ludger Paschen & Matthew Stave (eds.), *Language documentation reference corpus (DoReCo) 1.0*. Berlin & Lyon: Leibniz-Zentrum Allgemeine Sprachwissenschaft & laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2).

Witzlack-Makarevich, Alena, Saudah Namyalo, Anatol Kiriggwajjo & Zarina Molochieva. 2022. Ruuli DoReCo dataset. In Frank Seifart, Ludger Paschen & Matthew Stave (eds.), *Language documentation reference corpus (DoReCo) 1.0*. Berlin & Lyon: Leibniz-Zentrum Allgemeine Sprachwissenschaft & laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2).

Włodarczak, Marcin & Mattias Heldner. 2017. Respiratory constraints in verbal and non-verbal communication. *Frontiers in Psychology* 8. 1–11.