COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 49 (2025) e70087

ISSN: 1551-6709 online
DOI: 10.1111/cogs.70087

How Does Speaking A Free Word Order Language Influence Sentence Planning and Production? Evidence From Pitjantjatjara (Pama-Nyungan, Australia)

Evan Kidd,[a,b,c] Gabriela Garrido Rodríguez,[b,c] Sasha Wilmoth,[b,d] Javier E. Garrido Guillén,[e] Rachel Nordlinger[b,d]

[a]*School of Literature, Languages, and Linguistics, The Australian National University*
[b]*ARC Centre of Excellence for the Dynamics of Language*
[c]*Language Development Department, Max Planck Institute for Psycholinguistics*
[d]*School of Language and Linguistics, The University of Melbourne*
[e]*School of Engineering and Science, Tecnológico de Monterrey*

Received 18 March 2023; received in revised form 31 May 2025; accepted 9 June 2025

Abstract

Sentence production is a stage-like process of mapping a conceptual representation to the linear speech signal via grammatical rules. While the typological diversity of languages is vast and thus must necessarily influence sentence production, psycholinguistic studies of diverse languages are comparatively rare. Here, we present data from a sentence planning and production study in Pitjantjatjara, an Australian Indigenous language that has highly flexible word order. Forty-nine ($N = 49$) native speakers described pictures of two-participant scenes while their eye-movements were recorded. Participants produced all possible orders of agent, patient, and verb. There was a general preference to produce agent-initial orders, but word order was influenced by the semantic properties of agent and patient referents (± human). Analyses of participants' eye-movements revealed early *relational encoding* of the entire event, whereby speakers distributed their attention between agent and patient referents in a manner that is different than typically observed in languages that have more restricted word order options. Relational encoding was influenced by the word order that participants eventually produced. The results provide evidence to suggest that sentence planning in Pit-

Correspondence should be sent to Evan Kidd, School of Literature, Languages, and Linguistics, The Australian National University, Baldessin Precinct Building, Acton, Canberra, ACT 2601, Australia. E-mail: evan.kidd@anu.edu.au

jantjatjara is a hierarchical process, in which early relational encoding creates a wholistic conceptualization of an event, possibly driven by pressure to decide upon one of many possible word orders.

Keywords: Sentence planning; Relational encoding; Eye-tracking; Production; Free word order; Australian languages

1. Introduction

Producing a sentence involves the complex process of mapping a conceptual representation to a linear speech signal via language-specific grammatical rules (Bock & Levelt, 1994; Levelt, 1989). An obvious between-language variable that is likely to affect this process is the grammar of a language. There is large typological variation in even the most basic grammatical constructs (Skirgård et al., 2023), and yet, psycholinguistics has rarely studied production processes outside of European languages. For instance, Jaeger and Norcliffe (2009) estimated that there are published sentence production studies in less than 1% of the world's 7000 languages, with most research in the broader domain of psycholinguistics focusing on English and other larger Germanic and Romance languages spoken in Western Europe (Berghoff & Bylund, 2025; Collart, 2024; Kidd & Garcia, 2022). This replicates a widespread sampling problem in cognitive science (see Blasi, Henrich, Adamou, Kemmerer, & Majid, 2022; Henrich, Heins, & Norenzayan, 2010). Against the backdrop of poor language coverage, there have been notable efforts to study sentence production in typologically diverse languages (e.g., Christianson & Ferreira, 2005; Egurtzegi et al., 2022; Hwang & Kaiser, 2014a; Myachykov, Garrod, & Scheepers, 2010; Norcliffe & Jaeger, 2016; Norcliffe, Konopka, Brown, & Levinson, 2015; Nordlinger, Garrido Rodriguez, & Kidd, 2022; Santesteban, Pickering, Laka, & Branigan, 2015; Sauppe, Norcliffe, Konopka, Van Valin, & Levinson, 2013; Tanaka, Branigan, McLean, & Pickering, 2011). In the current paper, we add to this literature by presenting the results of a study on sentence planning and production in Pitjantjatjara, an Indigenous Australian language with typological features rarely explored in psycholinguistic studies. Notably, we show that one prominent feature of Australian Indigenous languages—their ability to freely order sentence elements without affecting message meaning—strongly influences sentence planning.

1.1. Sentence planning and production

Sentence production is an incremental and stage-like process, involving *conceptualization*, *linguistic encoding*, and *articulation* (Bock & Levelt, 1994; Ferreira & Slevc, 2007; Garret, 1975; Levelt, 1989; Roelofs & Ferreira, 2019). While there is broad agreement about the procedural nature of sentence production, the degree to which the stages operate in parallel and thus the size of the units they operate over has been subject to debate. On the one hand, production could be radically incremental (i.e., *linear*), operating over single concepts/units. Thus, to describe a scene, a speaker may simply identify a prominent entity and choose that concept as a starting point (Gleitman, January, Nappa, & Trueswell, 2007). On the other hand, the speaker could derive a more elaborate event representation prior to linguistic encoding (i.e., *hierarchical*), which lays a foundation for lexical selection and structural assembly (Bock, Irwin, & Davidson, 2004; Griffin & Bock, 2000; Konopka & Brown-Schmidt, 2014). Rather than being in opposition to each other, the weight of evidence is consistent with the idea that speakers can flexibly adapt the size of the planning unit across contexts and tasks. However, a growing body of research on typologically diverse languages has revealed a moderating role for the grammar of a language in sentence planning, as revealed through eye-tracking during picture description.

Pioneered by Griffin and Bock (2000), eye-tracking is revealing about sentence planning in several ways. First, the pattern of eye-movements in the message encoding stage (circa 0–600 ms post picture onset), typically termed *event apprehension* because the picture description method requires participants to gain a construal of the event, is revealing about the degree to which speakers engage in linear versus hierarchical incremental planning. Second, the patterns of eye-movements after event apprehension, during the *linguistic encoding* stage (from circa 600 ms up until speech onset), reveal the linearization of the linguistic signal. During linguistic encoding, speakers fixate on referents in the order in which they will eventually produce them in their picture description. Thus, the eye-tracking during picture description methodology can provide key insights into the stage-like incremental production processes in language (Levelt, Roelofs, & Meyer, 1999).

Evidence demonstrating incremental linear planning during event apprehension frequently comes from work on languages like English and Dutch, which have very little morphological case and relatively fixed word orders. In an eye-tracking study with English-speaking participants, Gleitman et al. (2007) showed that subliminal attentional cues to pictures in a scene increased the likelihood that it would be produced first, even when doing so resulted in the production of a more complex structure (e.g., a passive). Schlenter, Esaulova, Dolscheid, and Penke (2022) reported that referential cues also guide sentence formulation in German, which has morphological case and some flexibility in word order. Thus, according to the linear incremental approach, speakers can access one conceptual component of an event and begin linguistic encoding from that starting point, flexibly encoding linguistic information downstream as they simultaneously build a more elaborate semantic representation of an event.

A key signature of hierarchical planning in eye-tracking studies of sentence production is *relational encoding*. During event apprehension, relational encoding is observed when speakers distribute their gaze between characters in a scene early in the planning process (Konopka, 2019; Konopka, Meyer, & Forest, 2018), under the assumption that this pattern of eye-movements indicates the creation of a conceptual representation of the entire event, which can then guide linguistic encoding. Accordingly, the eye-tracking record, on this account, should provide evidence of (at least partially) separable event apprehension and linguistic encoding processes (Norcliffe et al., 2015; Sauppe et al., 2013). In contrast, the absence of relational encoding typically results in long fixations to the first-mentioned character in the sentence, with attention directed at the second-mentioned character much later, often after speech onset. This latter pattern is indicative of linear planning.

1.2. Past eye-tracking literature in crosslinguistic perspective

Past research on sentence production using eye-tracking suggests that sentence planning is influenced by language-specific and language-independent factors (e.g., attention, referent accessibility). We focus here on research that has investigated how typological variables influence the planning and production of transitive clauses. A key finding in the past research on sentence planning across typologically diverse languages has been that planning can look very different from foundational results observed in English and Dutch. In particular, the speakers of typologically diverse languages may engage in greater amounts of relational encoding earlier in sentence planning. This occurs both when languages mark core argument roles more overtly and transparently, and, as we shall see, when a language has highly flexible word order.

Sauppe et al. (2013) conducted a sentence picture description experiment in Tagalog, a verb-initial Western Austronesian language spoken in the Philippines that has a typologically rare symmetrical voice system, in which voice marking on the verb (agent or undergoer voice) identifies either the agent or patient as a prominent syntactic argument (Himmelman, 2005), and which allows flexible ordering of core arguments (though this is voice-dependent, see Garcia, Roeser, & Kidd, 2023). They found that, while speakers of Tagalog predominantly focused on the agent character during event apprehension, the relative degree of attention to the patient character differed according to voice marking and word order, showing different patterns of relational encoding depending on these core properties of the language.

Norcliffe et al. (2015) found a similar pattern in Tseltal (Mayan, Mexico), another canonically verb-initial language with the typologically rare VOS word order that has significant argument marking on the verb, but which also permits SVO word order. In a comparison with Dutch, they found that sentence planning for SVO sentences looked similar across both languages and was consistent with linear planning. However, Tseltal speakers' planning for their canonical VOS word order was very different, with comparatively greater attention to the patient character during event apprehension than in SVO sentences (the same tendency occurring when these active word orders were passivized). The general conclusion from both Tseltal and Tagalog is that producing verb-initial sentences requires a greater degree of relational encoding earlier in sentence planning. This is, to some degree, to be expected: verb selection requires identification of its argument roles, and given that both Tagalog and Tseltal contain obligatory morphological processes on the verb that link to core arguments, it stands to reason attention to both the agent and patient is important in the early stages of planning.

Evidence for earlier relational encoding and, therefore, hierarchical planning does not come solely from verb-initial languages. In a series of studies comparing Korean (Koreanic, Korea) and English (Hwang & Kaiser, 2014a, 2014b, 2015), English speakers prioritized finding a starting point to begin a sentence (i.e., most often an agent) and its relationship to the action to be described, whereas Korean speakers, whose language is canonically SOV and has case marking on nouns, prioritized relational encoding between the agent and patient before the linguistic encoding of the verb.

Similar processes have been argued to occur in Hindi (Indo-Aryan, India) and Basque (isolate, Spain). Hindi and Basque are split-ergative languages (Dixon, 1994), where in some circumstances (differing, depending on the language), transitive and intransitive subjects are treated as distinct, the latter being treated the same as transitive objects (*ergative-absolutive* alignment). This results in a different pattern of alignment than in nominative-accusative languages, which treat transitive and intransitive subjects as the same. Sauppe et al. (2021) asked Hindi speakers to describe transitive scenes using either present tense, which requires nominative-accusative alignment, or perfective aspect, which requires ergative-absolutive alignment, while their eye-movements to event participants were recorded and their neural activity was measured using electroencephalography (EEG). The eye-movement data suggested that participants engaged in greater relational encoding when producing an ergative subject, as indicated by more distributed looks between the agent and the patient.

Egurtzegi et al. (2022) conducted a comparable study with speakers of Basque and Swiss German (i.e., they compared languages with ergative and nonergative case systems). They reported a later peak in fixations to the agent character in Basque compared to Swiss German 200–800 ms post picture onset, which they interpreted to indicate greater relational encoding. Note, however, that this interpretation of relational encoding differs from previous ones, since it was not the case that speakers of Basque and Swiss German differed in their distribution of attention to the agent in comparison to the patient (i.e., the later peak in Basque did not translate to fewer fixations to the agent and thus more to the patient across event apprehension). Indeed, a subsequent study comparing Basque and Spanish revealed a consistently *greater* focus on agents in Basque early in event description planning (Isasi-Isamendi et al., 2023). Interestingly, while the eye-movement data in both Sauppe et al. (2021) and Egurtzegi et al. (2022) were interpreted to indicate greater relational encoding in sentences containing ergative subjects, the EEG results showed a divergent pattern of results. Whereas ergative alignment in Basque was associated with increased theta- and alpha-band activity, Hindi showed a comparable pattern of results for sentences with unmarked nominative subjects. These differences were argued to reflect the specific grammatical condition governing subject choice in each language.

While many of the studies reviewed thus far involve languages with some degree of flexibility in word order, none have been claimed to have the word order freedom exhibited by the nonconfigurational languages of Australia (Austin, 2001, Hale, 1983, Nordlinger & Bresnan 2011). In the first study of an Australian language, Nordlinger et al. (2022) investigated sentence planning and production using eye-tracking in Murrinhpatha (non-Pama-Nyungan, Australia), a free word order polysynthetic Australian Indigenous language. Murrinhpatha is head marking, and so it is possible and very common for speakers to express all information about an event in a single verb, as in (1).

1. nungam-rirda
   3SGS.FEET(7).NFUT-kick
   "He kicked him" (Nordlinger et al., 2022, p. 191)

Overt noun phrases (NPs) can freely occur anywhere in the clause, without any clear change in the meaning of the sentence (Mujkic, 2013). Consistent with this fact, the participants in

Nordlinger et al. (2022) produced 10 out of the 11 possible word orders available to them, even though they were asked to describe scenes without any cueing. In the four most frequent word orders that contained two overt arguments (Agent-Verb-Patient [AVP], APV, PVA, PAV), Murrinhpatha speakers demonstrated a peak in looks to the agent within the first 300–400 ms after picture onset, followed by a peak in looks to the patient soon thereafter, a pattern of relational encoding that was more rapid than that observed in other languages, where such cross over in looks typically occurs in the linguistic encoding stage. There are two potential explanations for this rapid event conceptualization. One possibility is that it may be due to the pressure to assemble a complex polysynthetic verb that marks both subject and object roles. Nordlinger et al., however, suggest that the early relational encoding may be driven not by polysynthesis, but instead by the considerable number of word order options allowed in the language (for converging evidence from Korean, see Hwang & Kaiser, 2014b). Accordingly, early relational encoding may be due to the pressure to settle on the linear order of constituents to begin the process of linguistic encoding.

In the current study, we test these possibilities by adding another Australian language to the literature—Pitjantjatjara (Pama-Nyungan). Like Murrinhpatha, Pitjantjatjara has been described as having free word order (Bowe, 1990). For instance, a recent corpus study showed that speakers produced all possible variations of word orders in naturalistic speech except one (VOS), with no obvious dominant word order (S-initial: 41.81%, O-initial: 30.54%, V-initial: 27.65%, Wilmoth, Garrido Rodriguez, Nordlinger, & Kidd, 2025). However, unlike Murrinhpatha, it is not polysynthetic, instead marking participant roles on arguments via case marking. Thus, Pitjantjatjara provides an opportunity to separate the effects of word order flexibility from verbal argument marking and thus distinguish between the two possible explanations for the early relational encoding in the Murrinhpatha findings. In the next section, we provide a brief overview of the relevant linguistic features of Pitjantjatjara.

1.3. Pitjantjatjara overview

Pitjantjatjara is a variety of the Western Desert dialect chain, a group of Pama-Nyungan languages spoken over a vast, arid area of Central and Western Australia. It is primarily spoken in the A<u>n</u>angu Pitjantjatjara Yankunytjatjara Lands (APY Lands) in the north-west corner of South Australia, as well as neighboring areas in Western Australia and the Northern Territory (Fig. 1). It is spoken by over 3000 people, and is one of a small number of Australian languages that is learned as a first language by children (Marmion, Obata, & Troy, 2014). The language is the primary means of daily communication in many communities in the region. Our data was collected in Pukatja/Ernabella, the largest community in the APY Lands.

Like many Australian languages, Pitjantjatjara allows very high levels of word order flexibility. Some examples of different word orders in transitive sentences from our data are shown in (2) (where, ERG = ergative case, ACC = accusative case, PRS = present tense, PST = past tense, and IPFV = imperfective aspect). In these examples, and in the vast majority of our data, there is a straightforward mapping between grammatical function, thematic role, and case, whereby Subject = Agent = Ergative and Object = Patient = Accusative; minor exceptions to this are discussed below.

2. (a)

wati	*kutju-ngku*	*antipina*	*ila-ṉi*
man	one-ERG	fish(ACC)	pull-PRS
S		O	V

'A man is pulling a fish.' (20180926-PJ03)

(b)

minyma	*kutju-ngku*	*ila-ṉi*	*ngiṉṯaka*
woman	one-ERG	pull-PRS	monitor.lizard(ACC)
S		V	O

'A woman is pulling the lizard.' (20180926-PJ03)

(c)

puṉu	*kutju*	*wanangaṟa-ngku*	*waka-ṉu*
tree	one(ACC)	lightning-ERG	strike-PST
O		S	V

'Lightning hit a tree' (20180926-PJ03)

(d)

pintapinta	*ila-ṉingi*	*papa-ngku*
butterfly(ACC)	pull-PST.IPFV	dog-ERG
O	V	S

'The dog was pulling a butterfly.' (20180929-PJ09)

(e)

paḻtji-ni	*ngunytju-ngku*	*iṯi*	*kulupa*
wash-PRS	mother-ERG	baby	small(ACC)
V	S	O	

'The mother is washing the little baby.' (20181004-PJ24)

(f)

papa	*kutju-ngku*	*riitjamila-ṉi*
dog	one-ERG	chase-PRS
S		V

'A dog is chasing (a car).' (20180926-PJ03)

(g)

pikipiki	*kutja-ṉi*
pig(ACC)	cook-PRS
O	V

'(A man) is cooking a pig.' (20181007-PJ33)

(h)

witi-ni
V
catch-PRS

'(A woman) is catching (lizards).' (20181007-PJ35)

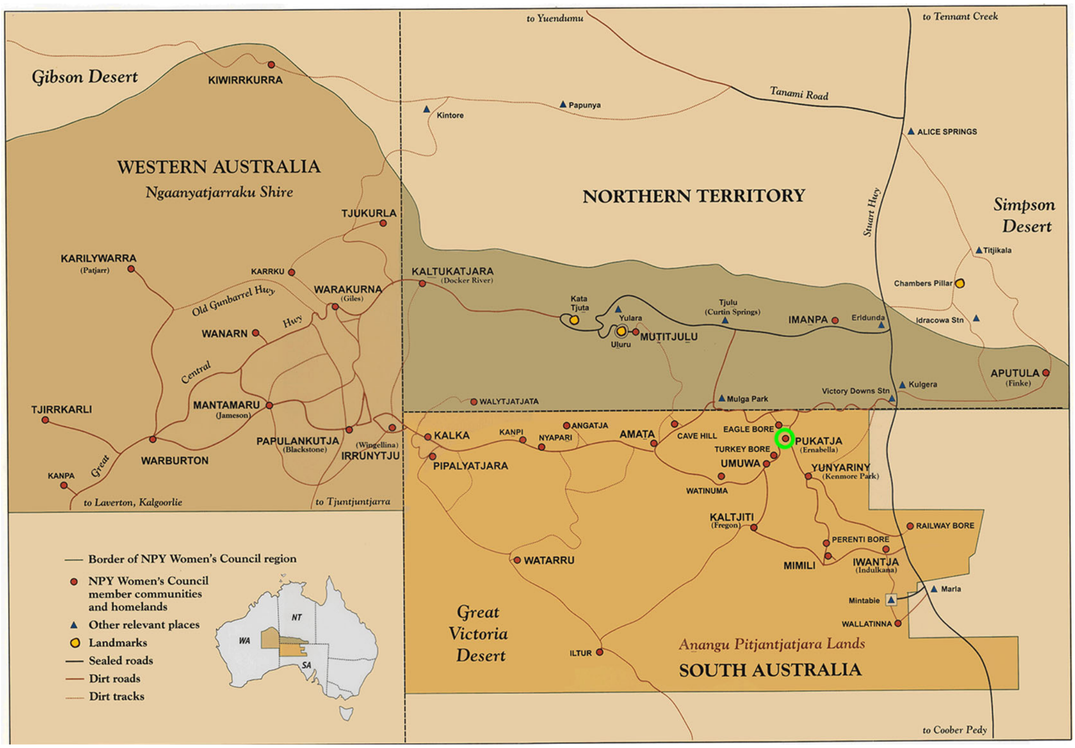

Fig. 1. Map showing the Ngaanyatjarra Pitjantjatjara Yankunytjatjara (NPY) region; the shaded region of South Australia is the APY Lands, where Pitjantjatjara is the primary language. This study was carried out in Pukatja/Ernabella, marked in green. *Source*: NPY Women's Council (https://www.npywc.org.au). Used with permission.

Like all Pama-Nyungan languages, Pitjantjatjara is dependent marking, meaning that it overtly marks participant roles on nouns. It has a split ergative case-marking system: pronouns are marked according to a nominative-accusative pattern (as in languages like English and German), while other nominals are marked according to an ergative-absolutive pattern. Common nouns and proper names have different forms of case markers. The paradigm of core case marking suffixes is shown in Table 1, presented according to a tripartite analysis (Goddard, 1982) where there are three core cases (ergative—*erg*, nominative—*nom*, accusative—*acc*).[1]

Case is marked on the right edge of the noun phrase, as seen with the ergative suffix in (2a, b, f). There is no marking of participants on verbs. Where noun phrases are expressed in our data set, they are almost always lexical noun phrases and thus marked according to an ergative-accusative pattern, as in the examples in (2). The majority of verbal lexemes in Pitjantjatjara are inherently transitive, intransitive, or ditransitive; there are also a small number of ambitransitive verbs such as *inkanyi* "play/perform" which may select a nominative subject argument, or an ergative subject and an accusative object. For example, the verb *riitjamila<u>n</u>i* "chase" is transitive; it has an ergative subject in (2f) despite no object being expressed. In

Table 1
Core case marking suffixes in Pitjantjatjara (Wilmoth, 2023)

		Ergative	nominative	accusative
Common nouns	Vowel-final	-ngku	Ø	Ø
	Consonant-final	-pangku ∼ -Tu	Ø	Ø
Proper names	Vowel-final	-lu	-nya	-nya
	Consonant-final	-Tu	-nga	-nga
Free pronouns	1sg	-lu	-lu	-nya
	3sg	-ṟu	-ṟu	-nya
	Other	Ø	Ø	-nya

(2g), the unmarked *pikipiki* "pig" is necessarily interpreted as the object of the transitive verb *kutjaṉi* "cook," and not as an intransitive subject.

1.4. The current study

In the current paper, we report on an eye-tracking study of sentence planning and production in Pitjantjatjara, using the same materials Nordlinger et al. (2022) used in their study of Murrinhpatha. Our primary goal was to determine the degree of relational encoding that occurs during sentence planning in the language and, thus, how extreme word order flexibility influences sentence planning. However, since there have been no experimental studies of Pitjantjatjara and because we used the same materials as were used in Murrinhpatha, we report several analyses pertaining to word order planning and production.

We first determined the range of word order variability in the language and what conceptual features of NP arguments might condition variation, investigating how conceptual accessibility of NP referents (i.e., agent and patient humanness) influenced word order selection. Conceptual accessibility of referents has long been known to influence grammatical function assignment and word order selection (e.g., Bock & Warren, 1985; Bock, Loebell, & Morey, 1992), and many studies have shown it to play a substantial role in determining word order choice in languages with flexible ordering options (e.g., Christianson & Ferreira 2005; Tanaka et al., 2011). A concept can be made more accessible through properties such as its discourse status (e.g., given vs. new information), its thematic role in the message (agent vs. patient, etc.), and its semantic properties such as humanness and/or animacy relative to other concepts in the message (Christianson & Ferreira, 2005). Previous work on languages with flexible word orders has shown that humanness plays a strong role in conceptual accessibility (Christianson & Ferreira 2005; Branigan et al., 2008; Esaulova et al., 2019; Norcliffe et al., 2015; Nordlinger et al., 2022; Tanaka et al., 2011), resulting in human referents tending to appear early in the sentence and/or be encoded as subject. In a comprehension study of NP omission in Odawa (Algonquian, North America), Christianson and Cho (2009) found the expectation of NP expression conformed to a series of feature hierarchies (e.g., agent > patient; topic > nontopic; human > less human) such that when the features are aligned (e.g., agent = topic = human), there is a greater expectation that the NP will be omitted. They also found an effect of relative accessibility when the hierarchies are misaligned: a topical agent is more likely to be overtly expressed when it is less human and the patient is human, for example. Thus, we hypothesized that in Pitjantjatjara, human referents would be more likely either to appear early in a sentence or be omitted altogether, thus exerting a significant influence on word order variation in the language, and that relative accessibility may play a role in the omission of NPs.

Second, we analyzed participants' eye-movements to agent and patient characters during event apprehension to investigate the degree to which sentence planning can be described as a linear or hierarchical process in Pitjantjatjara. We did so in two ways. First, following past studies (Norcliffe et al., 2015; Nordlinger et al., 2022), we analyzed whether participants' first fixation to a character had a significant influence on the word order they eventually produced. Our aim here was to determine whether encoding the first referent played some role in determining word order. If this is the case, it would constitute evidence in favor of linear planning. Second, we analyzed participants' fixation to event characters during event apprehension across the four most frequent word orders produced to determine the degree of relational encoding evident in the language. If the rapid relational encoding Nordlinger et al. (2022) observed in Murrinhpatha is due to speaking a free word order language, then we should also expect similarly rapid relational encoding in Pitjantjatjara; that is, we expect an early peak in looks to the agent followed by a rapid peak in looks to the patient, and that the pattern of looks would be influenced by the word orders participants' produced. Such a pattern would mean sentence planning in the language is also best described as a hierarchical process, and that early relational encoding observed in Murrinhpatha is a consequence of speaking a free word order language. However, Pitjantjatjara is typologically different from Murrinhpatha, since it obligatorily marks participant roles using case marking. It is possible that the pressure for early relational encoding in Murrinhpatha is due to it being a polysynthetic head-marking language, and would not be expected in Pitjantjatjara. In this case, while sentence planning in Pitjantjatjara might still be a hierarchical process, relational encoding may be less pronounced than in Murrinhpatha, and comparatively, more attention may instead be directed to the agent in the early stages of sentence planning, as observed in other ergative languages like Basque.

2. Method

2.1. Participants

Forty-nine adult first language speakers of Pitjantjatjara (30 females, *M*age = 36.81, *SD* = 11.98, range 18–60 years old) from the Indigenous community of Pukatja/Ernabella in the APY Lands of South Australia participated in the study. While Pitjantjatjara is the everyday language of the Pukatja community, multilingualism has long been the norm across Indigenous Australia (Vaughan & Singer 2018). Thus, many Pitjantjatjara speakers have varying degrees of competency in other Indigenous languages (including other Western Desert varieties, or other Pama-Nyungan languages such as Warlpiri or Arrernte, depending on family connections) and in English, with the latter acquired as a second language in school and used primarily in interactions with government services (e.g., education, health). Despite

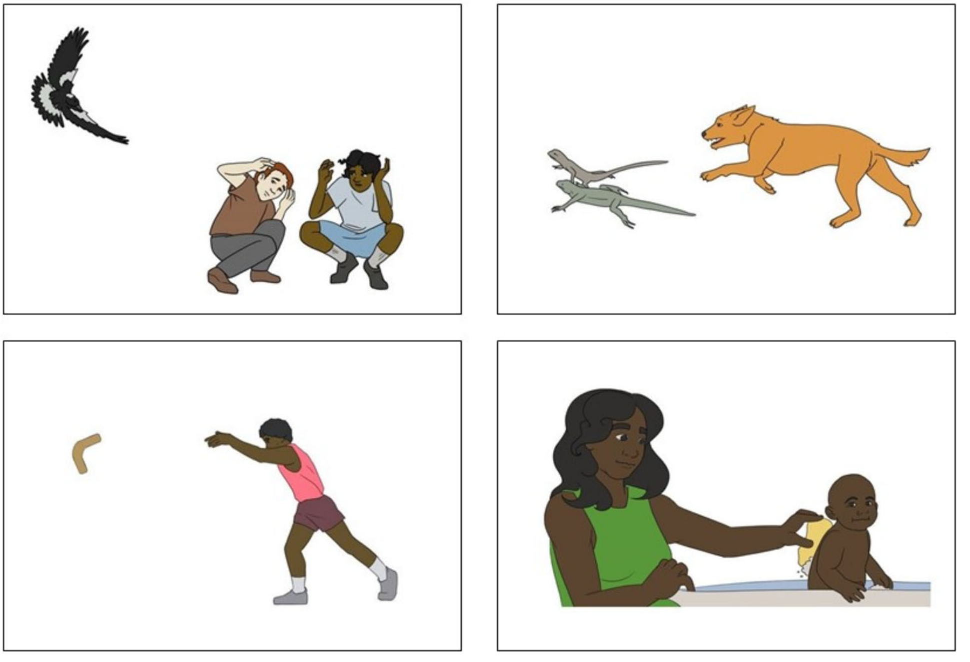

Fig. 2. Examples of stimulus items used in each condition.

this multilingualism, Pitjantjatjara was the dominant language for all participants. Participants gave written consent before the experiment began and were paid AUD$50 for their time. The study was approved by the University of Melbourne Human Ethics Committee (Project Ethics ID: 1237988).

2.2. Materials

We used the same picture description task as Nordlinger et al. (2022). Experimental pictures consisted of 48 images depicting transitive events (see Fig. 2). These pictures fully crossed thematic role (agent, patient) with humanness (human, nonhuman): 12 pictures had a human agent acting on a human patient (e.g., *girl pushes boy*), 12 had a human agent acting on a nonhuman patient (e.g., *man catches fish*), 12 had a nonhuman agent acting on a human patient (e.g., *crocodile chases children*), and 12 had a nonhuman agent acting on a nonhuman patient (e.g., *kangaroo boxes cow*). Twenty-nine of the pictures were also used in Norcliffe et al. (2015), with the remaining 19 created anew. All images depicted unique event characters that were known to Pitjantjatjara speakers and could be described using either existing words in the language or Pitjantjatjara borrowings from English (e.g., *putji* = cat, from "pussycat"). Two mirror-reversed versions of each experimental picture were constructed, such that the agent appeared on the left-hand side of the visual display and on the right-hand side in another display. An additional 96 pictures were served as fillers, which mostly depicted intransitive events (e.g., *baby crawling*, *tap dripping*).

Four blocks of stimuli were created containing the experimental pictures and the filler items (e.g., intransitive event pictures), with an additional four blocks created using the mirror image of test items (eight block in total, with participants only tested on four). Each block contained 36 items (12 tests, 24 fillers), and in total, participants saw 144 trials. The target pictures were arranged so that each target was followed by one or two filler pictures. Test images were pseudorandomized within blocks and the blocks were pseudorandomized across participants, resulting in 16 different test orders. The complete list of experimental items is provided in the study's Open Science Framework repository (https://osf.io/xesfv/). The stimuli were presented using SMI Experiment Centre software (version 3.7) in a Lenovo ThinkPad laptop computer with a 13'' monitor. Eye-movements were recorded using an SMI REDn Professional portable remote eye-tracker (SensoMotoric Instruments, Teltow, Brandenburg, Germany), with a sampling rate of 60 Hz (spatial resolution RMS 0.05°, tracking range 50×30 cm at 65 cm distance, distance to participant ∼ 57–60 cm), attached to the base of the laptop. Participants' picture descriptions were recorded using a Zoom H4n recorder.

2.3. Procedure

Participants were tested in a quiet room in a local house. They were told that they would see a series of pictures on a computer screen and would be asked to provide a short description of the event. Instructions were given mostly in Pitjantjatjara (except for loanwords such as "computer," and occasional code-switches to English) by the experimenter (the experimenter being a native English speaker and highly competent second language speaker of Pitjantjatjara). Participants were asked to sit in front of a portable eye-tracker and were told that both their eye movements and speech would be recorded. After this, participants provided written/verbal consent and continued with the experimental session.

The session began with seven practice trials to familiarize participants with the procedure. Participants were simply told to describe what they saw, such as "The girl is hopping on one foot." The eye-tracker was calibrated using a 9-point fixation dot followed by verification at the beginning of each block. The experimenter triggered each trial using an external keyboard. A trial began with a blank white screen that lasted 1 s. During this time, a small black dot appeared at the top of the screen, which could be either in the center or to the left or right of the screen. The dot was used to direct participants' attention away from the center of the screen to avoid random fixations on any of the test picture characters once the image was presented. After the 1 s presentation of the blank screen, a 130 ms high-pitched tone sounded before the image appeared on the screen. The experimenter sat at right angles to the participant, such that they were only present in the participants' peripheral vision. The experiment was presented across four different blocks to give participants the opportunity to rest if needed. The entire experimental session lasted approximately 30–40 min.

2.4. Sentence transcription and coding

Participants' descriptions of the test items were transcribed using ELAN[2] by the experimenter in conjunction with a native Pitjantjatjara speaker. Each transcribed sentence was then coded for word order (e.g., agent-patient-verb, or APV) by a single coder. The decision to code order based on thematic roles (AVP) rather than grammatical function (e.g., SVO) was made in order to account for responses such as (3), where the patient was expressed as a locative adjunct, rather than as an accusative object. There are no syntactic restrictions on order that apply to adjuncts such as this; like subjects and objects, they may appear in any position in the clause (LOC = locative, LOAN = transitive marker on loan verb, MV = medial verb).

(3)	*monkey-ngku*	*tjaata-ngka*	*painta-ma-n̠u*
	monkey-ERG	shirt-LOC	paint-LOAN-PST

'The monkey painted on the shirt.' (20181004-PJ26)

A small number of responses (0.68%) were included with a greater difference between the order of thematic role and order of grammatical functions, as shown in (4).

(4)	*kungka*	*nyangatja*	*nyina-nyi,*	*putji-ngku*	*miri-nu*	*mut̠i*
	girl(NOM)	this.one	sit-PRS	cat-ERG	scratch-PST	knee(ACC)

'This girl is sitting, a cat scratched (her) knee.' (20181004-PJ21)

This response contains two clauses: the first is an intransitive clause that introduces the human character (patient of the overall event, appearing here as nominative subject), while the second clause is a transitive clause describing the event as intended, with the girl's body part appearing as an accusative object following the verb. The introductory clauses in these examples employ the verbs *nyinanyi* "sitting" or *ngar̠anyi* "standing"; these verbs also have more grammaticalized senses in Pitjantjatjara and can be used in existential predicates, or as copulas (Goddard & Harkins, 2002). These clauses thus function to introduce the characters only, and do not describe separate, semantically rich events. Examples such as this were coded according to the first mention of the agent/patient; thus, (4) was coded as patient-agent-verb (PAV, with the relevant V being the transitive verb).

Some responses (0.34%) included a clause chain where one of the characters occurred in between the finite main verb and the nonfinite medial verb, as in (5). Clause chains are frequently used in Pitjantjatjara to describe complex events or sequences of events performed by the same subject (Goddard, 1988). Tense-aspect-mood is indicated on the final verb only, thus, we have coded examples such as this as APV, with the final verb as V. Typically, medial and main (final) verbs appear adjacent to one another in this type of clause chain (6), making examples such as (5) very uncommon. This construction *ngul̠ur̠a wat̠aparan̠i*, literally "scaring following," is conventionally used to mean "chase" in Pitjantjatjara (in variation with *riitjamilan̠i* "chase," based on the English "race"). Both *ngul̠u-* and *wat̠apara-* are transitive, and they share both core arguments in these examples.

(5)	*papa-ngku*	*ngulu-ra*	*mutuka*	*watapara-ni*
	dog-ERG	scare-MV	car(ACC)	follow-PRS
	'The dog is chasing the car'	(20180927-PJ04) (coded APV)		

Responses were excluded where it was clear the participant interpreted the event differently than intended. For example, in (7), the intended prompt was "the crocodile is biting the man"; however, the speaker reversed the agent and patient roles and interpreted the image as the man kicking the crocodile.

(6)	*papa-ngku*	*ngulu-ra*	*watapara-ni*	*mutuka*
	dog-ERG	scare-MV	follow-PRS	car(ACC)
	'The dog is chasing the car'	(20180927-PJ05) (coded AVP)		

2.5. *Data preprocessing*

2.5.1. *Production data*

There were a total number of 2352 possible observations (49 participants × 48 stimuli = 2352). Sentence onsets were manually segmented in ELAN. Trials were excluded if they were not transitive responses (403 observations), if speech onsets were 6500 ms or longer (47 sentences excluded), or if participants were not in the agent and patient roles as expected (24 sentences excluded). Thirteen trials were also excluded because of unclear recordings. The final data set for analysis consisted of 1865 sentences.

From this pool of 1865 possible observations, an additional 161 trials were lost due to technical problems,[3] resulting in 1704 trials. To preprocess the eye-tracking (ET) data, we excluded additional trials with track loss (i.e., where the eye-tracking software failed to detect eye-information [blinks, saccades, fixations]) or where trials lacked object-hit information (i.e., there was no record of looking at regions of interest: agent, patient, white space). Twenty-six trials (1.59% of trials) were excluded because of track loss, and 73 trials (4.28%) were excluded because of absent object-hit information.[4] The final data set for time-course analysis consisted of 1605 observations distributed across 11 different word orders: APV ($n =$ 809), AV ($n =$ 140), AVP ($n =$ 373), PAV ($n =$ 78), PV ($n =$ 96), PVA ($n =$ 57), V ($n =$ 32), VA ($n =$ 8), VAP ($n =$ 2), VP ($n =$ 9), VPA ($n =$ 1).

3. Results

3.1. *Preliminaries*

Our data and analysis scripts are available on the OSF (https://osf.io/xesfv/). Overall, we observed every possible word order in the data set (i.e., 11), the frequencies of which are shown in Table 2. Each participant produced, on average, 5.14 word orders ($SD =$ 1.59, range 1–8), and each stimulus item elicited, on average, 4.45 different word orders ($SD =$ 1.37,

Table 2
Frequency distribution of word orders produced

	Word order	Frequency	%
Agent-initial	APV	957	51.31
	AV	157	8.42
	AVP	415	22.25
Patient-initial	PAV	90	4.83
	PV	114	6.11
	PVA	76	4.08
Verb-initial	V	36	1.93
	VA	8	0.43
	VAP	2	0.11
	VP	9	0.48
	VPA	1	0.05
Total		1865	100.00

range 1–9). Agent-initial sentences were the most frequent orders ($n = 1529$), followed by patient-initial ($n = 280$) and verb-initial ($n = 56$).

The large number of word orders is comparable to Nordlinger et al.'s (2022) data from Murrinhpatha. Christianson and Ferreira (2005) reported similar variability in an offline sentence production study on Odawa. Thus, the languages show a large degree of word order flexibility that is not normally attested in the languages that are typically studied in psycholinguistic research, a result that is particularly striking because the characters depicted in all of the test pictures were unique (i.e., only occurrred once), and, therefore, the participants had no explicit discourse context that might induce greater flexibility through the impact of information structure (Simpson & Mushin, 2008). Thus, the large degree of word order variation exhibited in the experimental results reflects a syntactic flexibility in word order that cannot be attributed solely to information structure effects.

3.2. Conceptual accessibility

One context that likely explains some variation in word order is the conceptual accessibility of agent and patient referents. To address our first research question, we analyzed the influence of conceptual accessibility on word order in two ways. Specifically, we analyzed the influence of agent and patient humanness on selecting (1) Agent-before-Patient (APV and AVP = 1372) versus Patient-before-Agent (PAV and PVA = 166) order, and (2) an AV versus PV word order (see Table 3). We chose to distinguish the number of overt arguments in A- and P-initial productions because Nordlinger et al. (2022) found that humanness and thus conceptual accessibility interacted with NP omission, meaning that collapsing across 2-NP and 1-NP utterances may mask some effects. We did not analyze V-initial utterances because there were too few tokens.

We used Generalized Linear Mixed Models (Baayen, Davidson, & Bates, 2008; Barr, 2008; Jaeger, 2008) to assess the effect of agent and patient humanness and their interaction on word

Table 3
Frequency of A-initial and P-initial word orders by agent and patient humanness

Word order	Event description				
	H-H (%)	H-nonH (%)	nonH-H (%)	nonH-nonH (%)	Total (%)
2-NP					
APV	199 (20.79)	332 (34.69)	166 (17.35)	260 (27.17)	957 (100.00)
AVP	83 (20.00)	89 (21.45)	107 (25.78)	136 (32.77)	415 (100.00)
PAV	14 (15.56)	14 (15.56)	44 (48.89)	18 (20.00)	90 (100.00)
PVA	17 (22.37)	34 (44.74)	15 (19.74)	10 (13.16)	76 (100.00)
1-NP					
AV	61 (38.85)	9 (5.73)	60 (38.22)	27 (17.20)	157 (100.00)
PV	50 (43.86)	54 (47.37)	1 (0.88)	9 (7.89)	114 (100.00)
Total	424 (23.44)	532 (29.41)	393 (21.72)	460 (25.43)	1809 (100.00)

order choice. Generalized Linear Mixed Model (GLMMs) were estimated using the *glmer* function from the *lme4* package (version 1.1-21.1, Bates, Mächler, Bolker, & Walker, 2015) in R (version 4.1.2, R Core team, 2024), specifying a binomial likelihood with a logit link function. The dependent measure (i.e., word order) was defined as a dichotomous variable, differentiating between a specific word order produced (e.g., agent-initial sentences, coded as 1) compared to other sentence types (e.g., patient-initial sentences, coded as 0). The independent variables of agent and patient humanness were sum coded (+0.5 = human, −0.5 = nonhuman) (see Brehm & Alday, 2022; Schad, Vasishth, Hohenstein, & Kliegl, 2020). We entered each predictor separately in a series of stepwise selection procedures (evaluated via forward model comparison) (Pinheiro & Bates, 2000). We specified the maximal random effects structure that was justified by design and which allowed the models to converge (Barr, 2013; Barr, Levy, Scheepers, & Tily, 2013). Confidence intervals (95%) are provided for the regression coefficients to assess the significance of the effects.

3.2.1. *Agent-before-patient versus patient-before-agent*

The effect of agent humanness on word order choice was not significant ($\beta = 0.77$, $SE = 0.45$, 95% CI [−0.11, 1.66]). However, there was a significant effect of patient humanness ($\beta = -0.62$, $SE = 0.24$, 95% CI [−1.10, −0.14]), suggesting that Pitjantjatjara speakers preferred to produce agent-before-patient sentences when the patient was nonhuman. This effect was subsumed by a significant interaction between event characters ($\beta = 1.35$, $SE = 0.49$, 95% CI [0.39, 2.30]), which was driven by trials containing nonhuman agents. Specifically, speakers preferred to produce A-before-P sentences when all characters were nonhuman (A-before-P = 0.29 vs. P-before-A = 0.17) but preferred to produce P-before-A sentences in the Non-Human-Human condition (A-before-P = 0.2 vs. P-before-A = 0.36). The large standard deviation for the participants' slope for agent ($SD = 1.69$) suggests that the effect of the agent humanness varied across participants. The estimates for the best-fitting model are shown in Table 4.

Table 4
Estimates for the best-fitting models investigating the influence of agent and patient humanness on the production of (1) A-before-P versus P-before-A and (2) AV versus PV word orders

	Agent-before-Patient versus Patient-before-Agent			AV versus PV		
Variable	Regression coefficient (95% CI)	*SE*	*z*	Regression coefficient (95% CI)	*SE*	*z*
Fixed effects						
Intercept	2.92 (2.42, 3.43)	0.26	11.30***	1.84 (0.41, 3.27)	0.73	2.52*
Agent	0.77 (−0.11, 1.66)	0.45	1.71	−6.34 (−9.72, −2.95)	1.73	−3.67***
Patient	−0.62 (−1.10, −0.14)	0.24	−2.54*	3.53 (1.35, 5.71)	1.11	3.18**
Agent *Patient	1.35 (0.39, 2.30)	0.49	2.76**			
Random effects	Variance	*SD*		Variance	*SD*	
Participant						
Intercept	1.29	1.14		2.82	1.68	
Agent	2.86	1.69				
Item						
Intercept	0.46	0.68		8.23	2.87	

***$p < .001$; **$p < .01$; *$p < .05$.

3.2.2. Agent-only (AV) versus patient-only (PV)

Both fixed effects of character humanness were significant. Speakers were more likely to produce AV compared to PV sentences when the events featured nonhuman agents ($\beta = -6.34$, $SE = 1.73$, 95% CI [−9.72, −2.95]) and events depicting human patients ($\beta = 3.53$, $SE = 1.11$, 95% CI [1.35, 5.71]). The large SD for items intercept ($SD = 2.87$) suggests that items vary strongly in how likely they are to elicit a single argument response. The estimates for the best-fitting model are shown in Table 4.

3.3. Perceptual accessibility

Our goal in the perceptual accessibility analysis was to determine whether a first *sustained* fixation on either the agent or patient referent was significantly related to the word order that was eventually produced. By "sustained," we mean that it would be long enough to allow lexical access of the referent, thereby beginning the linguistic encoding process. Thus, we investigated whether first fixating on an agent resulted in an agent-initial word order, and whether first fixating on a patient resulted in a patient-initial word order, as occurs during linear planning (e.g., Gleitman et al., 2007; Schlenter et al., 2022). We defined first fixations as those fixations that occurred for at least 100 ms *after* the first saccade following picture onset, within a 600 ms time window. This time window is slightly longer than past studies that have used similar definitions of first fixation (e.g., 400 ms, as used by Gerwien & Flecken, 2016; Nordlinger et al., 2022); however, using a longer window enabled us to include more trials (and results did not differ across a more restricted definition of the time window). Table 5 shows the distribution of first fixations for agent- and patient-initial word orders. Note that

Table 5
Agent- and patient-initial word orders as a function of first fixated character

	First fixations			
Word order	Agent (%)	Patient (%)	White space (%)	Total (%)
Agent-initial	211 (43.69)	167 (34.58)	105 (21.74)	483 (100.00)
Patient-initial	32 (37.21)	33 (38.37)	21 (24.42)	86 (100.00)
Total	243 (42.71)	200 (35.15)	126 (22.14)	569 (100.00)

Table 6
Estimated regression coefficients for the model comparing agent-initial versus patient-initial sentences as a function of first fixation (agent vs. patient) and humanness of the agent and patient (human vs. nonhuman)

	Agent-initial versus Patient-initial		
Variable	Regression coefficient (95% CI)	*SE*	*z*
Fixed effects			
Intercept	2.05 (1.59, 2.50)	0.23	8.80***
First fixation	0.25 (−0.29, 0.78)	0.27	0.90
Agent	−0.23 (−1.06, 0.60)	0.42	−0.54
Patient	−0.22 (−0.73, 0.30)	0.26	−0.84
Random effects	Variance	*SD*	
Participant			
Intercept	0.44	0.66	
Agent	2.25	1.50	
Item			
Intercept	0.03	0.17	

***$p < .001$.

the number of observations in this data set is smaller than the overall number of observations because many first fixations fell outside of our definition.

Table 5 shows that a greater percentage of first fixations fell on the agent than the patient and white space when agent-initial word orders were produced, although the distribution of first fixations to the agent and patient were similar when patient-initial word orders were produced. The data were analyzed using Generalized Linear Mixed Models, estimated using the glmer() function in the *lme4* package (version 1.1-27.1, Bates et al., 2015) in R. All models predicted word order type (i.e., categorical dependent variable), where we modeled the success of producing: Agent-initial versus Patient-initial utterances as a function of first fixated character (i.e., categorical independent variable sum-coded as agent = +0.5, patient = −0.5), and the interactions between first fixated character and character humanness. Table 6 shows the results of the best-fitting model. As observed in several other studies (Norcliffe et al., 2015; Sauppe et al., 2013), the participants' first fixation was not systematically related to word order choice ($\beta = 0.25$, $SE = 0.27$, 95% CI [−0.29, 0.78]).

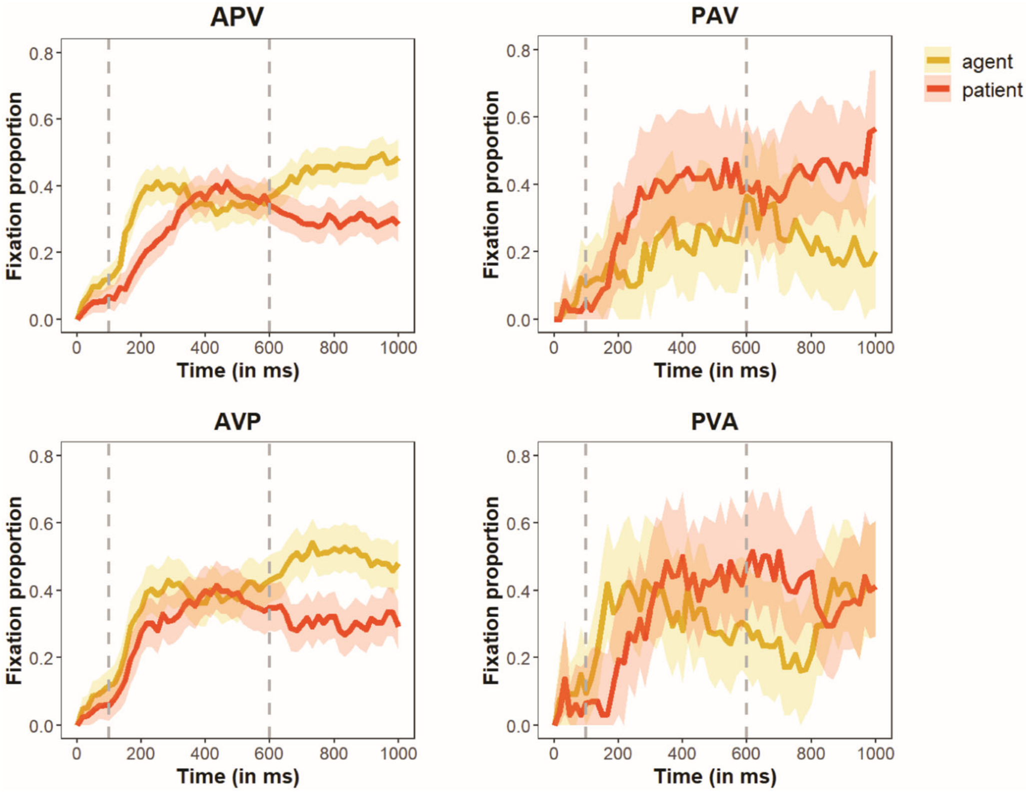

Fig. 3. Proportion of agent- and patient-directed fixations in agent-initial and patient-initial utterances in Pitjantjatjara. Time zero corresponds to picture onset. Ribbons indicate 95% CIs, calculated for each sampling step. Dashed lines indicate event apprehension time window.

3.4. *Patterns of eye-movements during event apprehension*

In our final set of analyses, we investigated the time-course of fixations to the agent and patient characters during event apprehension, starting from 100 to 600 ms post-picture onset. Our goal was to analyze the eye-movement data to determine the degree of relational coding that exists in Pitjantjatjara, as indexed by the distribution of fixations to the agent and patient referents, and whether relational encoding varies across word order.[5] Following Nordlinger et al. (2022), we analyzed the four most frequent word orders that contained two overt NPs: APV, AVP, PAV, and PVA.

Fig. 3 shows the time-course graph of the proportion of fixations to the agent and patient characters from picture onset until 1000 ms for APV, AVP, PVA, and PAV sentences. There are two notable features of the eye-movement records. First, there appears to be degrees of relational encoding between agent and patient characters in all word orders during the event apprehension window (100–600 ms). Notably, there are initial looks to the agent in all word orders (though considerably less in the PAV order), and then subsequent looks to the patient character, a pattern that seems to vary across word orders.

We used a Bayesian Generalized Additive Mixed Models (GAMMs, Hastie & Tibshirani, 1990; Porretta et al., 2018; Wood, 2017; Zuur, Ieno, Walker, Saveliev, & Smith, 2009; Fahrmeir & Kneib, 2011) to analyze the time course of fixations to event characters during event apprehension. Mixed models represent an appropriate way to model repeated measures, and the major advantage of GAMMs is that they estimate nonlinear (smoothed) relationships between predictors (categorical or continuous) and dependent variables, thus allowing us to model the nonlinear changes in fixation patterns across time. Our choice of a Bayesian approach was largely pragmatic, since it allowed us to run multinomial mixed models, analyzing the relative looks to agent and patient (and white space) in one model (see below). For our purposes, the technique allowed us to determine whether there was relational encoding during event apprehension (i.e., distributed looks to the agent and patient) and whether this varied according to the word order produced. In addition, the use of a Bayesian framework allows proper probability statements (i.e., allows a direct judgment of the probability of fixating a specific character) and responds to the fact that both random effects and penalized splines (i.e., smooth terms) lie under the same theoretical framework, in which Bayesian inference is natural. Another advantage is that we can easily obtain marginal posterior mean estimates of the probability of fixating each character as a function of word order.

Our dependent variable was categorical with three levels (agent, patient, and white space), and thus we modeled the probability of fixating the agent, the patient, and white space using a multinomial (categorical) probability distribution. The model of interest included word order (four levels: APV, AVP, PVA, PAV) and smooth interactions for word order and time (continuous variable), as well as smooth terms for time by participants and items. This model was compared to a null model without these terms. We used a logit link function for both the probability of fixating the agent and the probability of fixating the patient, compared to fixations on white space (baseline category). The analyses were conducted in R (version 4.4.1; R Core Team, 2024) using the brm() function in the *brms* package (version 2.21.0, Bürkner, 2017, 2018).

We used default priors provided by the *brms* package (i.e., flat improper priors for all parameters and noncentral Student's *t* priors for all variances). These noninformative priors represent the fact that we had no prior knowledge of the relationship between fixations and word order over time in the language. We simulated 10,000 Markov Chain Monte Carlo iterations after a burn-in period (i.e., warm-up draws that might not have yet converged to the posterior distribution) of 10,000 draws for each of the four chains, resulting in a total of 40,000 posterior draws.

Model comparison was carried out using the Widely Applicable Information Criterion (WAIC), a fully Bayesian approach that considers both the model fit and complexity. The smaller the WAIC, the better the model under consideration. We compared two models: The first model (m1) included the standardized values (z-transformed) of the size of the areas of interest (agent and patient characters) as nuisance predictors (Sassenhagen & Alday, 2016), the smooth term of time, and smooth terms for time by participants and items. The second model (m2) had the same structure but included the effect of word order and smooth inter-

Table 7
Results of the Bayesian multinomial GAMM modeling the probability of fixations to the agent, patient, and white space (baseline) as a function of word order (APV-reference, PAV, AVP, and PVA), time, and agent/patient size (AOI), during event apprehension (100−600 ms)

Parameter	Estimate (posterior mean)	*SE*	Credible interval (95%)	Rhat	Bulk ESS
Fixations to the Agent(vs. white space)					
Intercept	0.61	0.30	0.02, 1.20	1.00	4783
PAV	−0.54	0.22	−0.97, −0.12	1.00	43,076
AVP	−0.27	0.10	−0.46, −0.08	1.00	35,772
PVA	−1.55	0.17	−1.88, −1.21	1.00	40,024
Agent AOI-size	0.72	0.17	0.38, 1.07	1.00	9316
Patient AOI-size	−0.49	0.39	−1.26, 0.29	1.00	7776
Fixations to the Patient (vs. white space)					
Intercept	−0.00	0.30	−0.60, 0.58	1.00	4636
PAV	0.54	0.19	0.17, 0.91	1.00	43,661
AVP	0.28	0.10	0.09, 0.49	1.00	35,438
PVA	0.01	0.17	−0.33, 0.36	1.00	37,522
Agent AOI-size	0.24	0.17	−0.08, 0.57	1.00	11,013
Patient AOI-size	1.08	0.36	0.37, 1.78	1.00	9183

actions for word order (four levels: APV, AVP, PVA, PAV) and time (continuous variable) as predictors. That is, we tested whether the inclusion of word order and its interactions with time better explained the data (i.e., whether m2 fit the data better than m1). This is what we found: including word order and the smooth interactions of word order and time improved the model fit compared to the model without those predictors ($\text{WAIC}_{\text{m1}} = 16{,}194.5$, $SE = 152.4$; $\text{WAIC}_{\text{m2}} = 15{,}982.9$, $SE = 153.8$).

In addition, we calculated the expected log pointwise predicted density (elpd), which is a predictive accuracy measure of both models. A description of the procedure can be found in our OSF documentation. For the null model (m1), we obtained an $\text{elpd}^{\text{A}}_{\text{loo}} = -8101.15$ (se ($\text{elpd}^{\text{A}}_{\text{loo}}$) $= 76.23$) and for the proposed model with word order (m2), we obtained an $\text{elpd}^{\text{B}}_{\text{loo}} = -7995.97$ (se ($\text{elpd}^{\text{B}}_{\text{loo}}$) $= 76.92$). The difference between the two models is -105.18(se ($\text{elpd}^{\text{A}}_{\text{loo}} - \text{elpd}^{\text{B}}_{\text{loo}}$) $= 16.03$), meaning that the difference in elpd is much larger than several times the estimated standard error of the difference, indicating that the model with word order and its interaction with time is expected to have better predictive performance than the null model.

We assessed convergence of our preferred model (i.e., m2) through trace plots and the $\hat{R}$ statistic (R hat), which showed a value of 1.0 for all parameters (i.e., the sampler converged). In addition, we obtained a minimum bulk effective sample size (BULK ESS) of 4000. An effective sample size greater than 1000 is sufficient for stable estimates (Bürkner, 2017). Thus, the results suggest that we have reliable estimates. Table 7 presents the posterior summary, reporting: the posterior mean for each of the parameters in the population-level part model; the posterior standard deviation (i.e., the standard error); the 95% symmetric credible intervals; and finally, the posterior standard deviation estimates for the spline coefficients are presented.

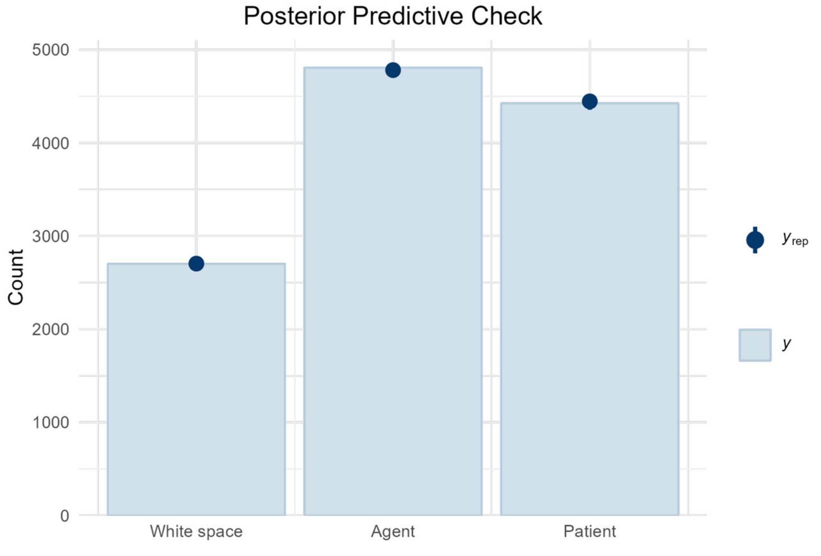

Fig. 4. Posterior predictive check for the best-fitting model (including word order and its interaction with time). The plot displays observed fixation counts as bars (i.e., y is the observed data) and the predicted counts from posterior simulations as overlaid dots (i.e., Yrep represents the simulated data). The close alignment between the observed and simulated data suggests excellent fit to the data.

To check the goodness of fit of the model, we performed posterior predictive checks. The idea is that simulated data generated with the posterior predictive distribution should resemble the observed data. Given that our target variable is categorical, we sampled from a multinomial (categorical) distribution. The only summary statistic that we can meaningfully visualize is the number of simulations falling into each category. What we compared, then, was the number of fixations on each character (and white space) in the observed data versus the number of fixations predicted by the model through simulation. As shown in Fig. 4, the distributions are almost identical, indicating that the model replicated the observed data with high accuracy. Thus, we can have confidence in the accuracy of the model predictions.

We interpret the fact that our full model was a better and accurate fit to the data as evidence that, in Pitjantjatjara, word order affects the probability of fixating to the agent versus the patient (and white space) during event apprehension. We next want to determine the nature of this planning. A well-known trade-off in statistics is that, as the flexibility of the model increases (as in Bayesian multinomial GAMMs), the less interpretable it becomes (James, Witten, Hastie, Tibshirani, & Taylor, 2023). Thus, rather than interpreting the posterior estimates of the model coefficients (Table 7), we used the posterior draws to visualize the posterior probability of fixating the agent, the patient, and white space as a function of word order and time, to describe their effect on the fixation behavior (i.e., to determine the pattern of relational encoding). Fig. 5 shows the posterior probability of fixation for the different word order configurations (panel A), indicating that there was a higher probability of fixating the patient, relative compared to the agent and white space, across all word orders, although the

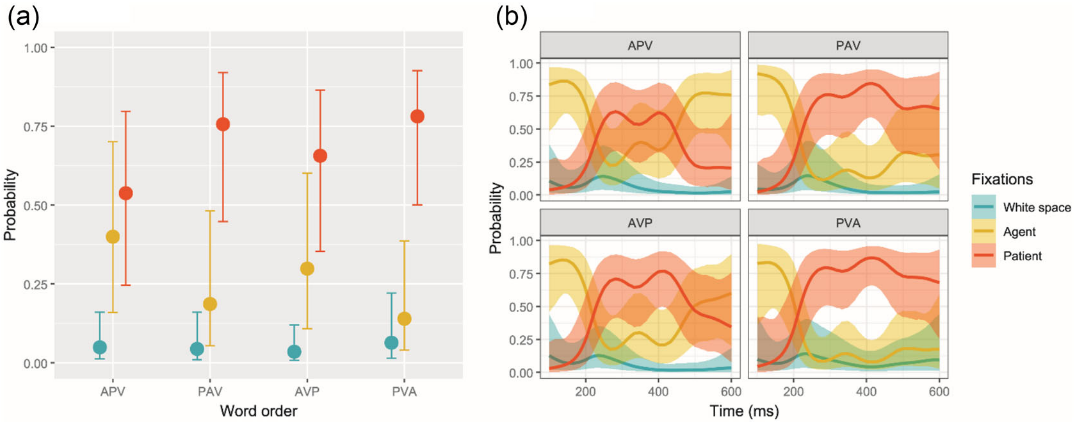

Fig. 5. Panel A shows posterior mean probabilities of fixating the agent (yellow lines), the patient (orange lines), and white space (blue lines) for each word order along with their 95% credible interval. Panel B shows posterior mean probabilities of fixating the agent, the patient, and white space over time (i.e., during event apprehension) for each word order. Shaded areas indicate 95% pointwise credible intervals.

difference differed in magnitude across word orders. Panel B depicts the posterior probability of fixating the event characters and white space for each word order configuration across time. It shows that the probability of fixating the agent was higher across all word orders in the first 200 ms after picture onset. After this time, there is a switch to fixating the patient in all word order types, indicating gaze alternations between the agent and patient characters during the first 600 ms of picture onset; that is, as Nordlinger et al. (2022) Murrinhpatha data, we have evidence of early relational coding within event apprehension. In agent-initial word orders, fixations to the patient drop again around 450–500 ms, after which the probability of fixating the agent increases again toward the end of the time window. In patient-initial word orders, the probability of fixating the patient is higher after 200 ms until the end of the time window (600 ms). We have added posterior probability difference plots to the Appendix, which show the probability of fixating the agent in pairwise word order comparisons for readers interested in exploring word order differences further.

4. Discussion

In the current study, we investigated sentence planning and production in Pitjantjatjara, a Pama-Nyungan language spoken in central Australia. As is common in this language family, Pitjantjatjara has highly flexible word order and an ergative case system. Thus, it is unlike many of the languages typically studied in psycholinguistics (Berghoff & Bylund, 2025; Collart, 2024; Jaeger & Norcliffe, 2009). We had two research aims. The first was to investigate the determinants of word order variation in Pitjantjatjara, with a focus on how the conceptual accessibility of referents conditions variation. The second, which addressed our main research question, was to investigate online sentence planning in the language using eye-tracking, where we aimed to determine how the language's flexible word order impacted the earliest stage of planning. We discuss each of these aims and the results that bear upon them in turn.

4.1. Word order variation in Pitjantjatjara

Concerning our first aim, we found that Pitjantjatjara has highly flexible word order. Despite participants having no instructions other than to simply describe one- and two-participant scenes, we observed every possible word order available to speakers in the language (i.e., 11). This is consistent with similar past work by Nordlinger et al. (2022) in Murrinhpatha and Christianson and Ferreira (2005) in Odawa. It is worth pointing out the difference between these languages and other well-studied languages like German, which has some flexibility of word order, but not to the same degree as found in languages such as Pitjantjatjara, Murrinhpatha, and Odawa. In similar studies to ours, Sauppe (2017) reports only three word orders (SVO, S-aux-O-V, and the passive) with clearly dominant use of one word order (SVO, accounting for >75% of responses), and Schlenter et al. (2022) report only two word orders (SVO actives and passives). Thus, languages like Pitjantjatjara ramp up the degree of flexibility in ways that are rarely studied in the field.

While all possible word orders were available to the speakers, the different options were not evenly distributed, and the word order choice was significantly influenced by agent and patient humanness. In our analysis of sentences containing two overt NPs, we found that A-initial word orders were more common when patients were nonhuman, which is consistent with the common observation that object arguments tend to be lower on the animacy hierarchy (Aissen, 1997; Silverstein, 1976) and that word order choice in languages with high flexibility may be driven by the relative accessibility of referents (Christianson & Ferreira, 2005; Christianson & Cho, 2009). Overall, we found that human arguments tended to be dropped: in our analysis of one argument descriptions (i.e., AV vs. PV), we found that participants were more likely to produce AV descriptions when there were nonhuman agents and human patients, with the opposite being the case for PV descriptions. Indeed, the greatest number of single argument sentences were in the Human−Human condition, where participants dropped the agent and patient in almost equal measures. This is consistent with the findings of Christianson and Cho (2009) for Odawa, and attests to the general conceptual prominence of human referents in shaping production, providing evidence in favor of our first hypothesis.

The predominance of A-initial descriptions in the data is different from the distribution of word orders Nordlinger et al. (2022) observed for Murrinhpatha, which had more P-initial and V-initial word order descriptions, although there are many similarities in the pattern of results. Like in Pitjantjatjara, A-before-P word orders in Murrinhpatha were more likely to have nonhuman patients, and human referents were more likely to be omitted in descriptions containing one overt NP. Additionally, one notable context elicited a greater number of P-initial word orders in both languages—when pictures had nonhuman agents acting upon human patients (e.g., *the crocodile chasing the children*). These are contexts that typically elicit passive sentences in European languages (e.g., Bock et al., 1992) or patient-first sentences in other languages that allow the flexible ordering of arguments (e.g., Japanese, Tanaka et al., 2011), again attesting to the prominence of human referents.

It is important to note, however, that the effect of conceptual accessibility on word order was probabilistic and does not explain the full range of the variation in the data. Pitjantjatjara speakers have many ordering options available to them, which is no doubt influenced by several factors for which we have not accounted (e.g., information structure, Simpson & Mushin, 2008). Importantly, the range of word orders that we observed in all likelihood *underestimates* the level of word order flexibility in the language. In an analysis of corpus data, Wilmoth et al. (2025) show that the APV order, which we found accounted for around half of our observations, is comparably rare in naturalistic speech, being used in only 16.28% of all clauses. In Wilmoth et al.'s corpus data, all word orders except VPA were attested (a likely gap in the data, given that it occurred in our experimental data), with the most frequent word orders being PV (20.93%) and V (20.21%). This means that our finding of a higher rate of A-initial word orders may be a byproduct of the experimental method. In particular, it is possible that the picture description method, which presents new referents in each picture, leads to an inflated number of A-initial utterances. Significantly more empirical and theoretical work is required to understand the determinants of word order variation in Pitjantjatjara and other Australian languages.

4.2. Sentence planning in Pitjantjatjara

Our eye-movement analysis revealed evidence of a pattern of early relational encoding that is very similar to that observed by Nordlinger et al. (2022) for Murrinhpatha: within event apprehension and for all word orders, we observed an initial early peak in looks to the agent referent, followed by rapid looks to the patient. Thus, like Murrinhpatha speakers, speakers of Pitjantjatjara engage in early relational encoding. This is different from other languages that have ergative marking, such as Hindi and Basque, where relational encoding has been inferred from comparatively fewer looks to the agent (Hindi, Sauppe et al., 2021) or from a later peak in looks to the agent (Basque, Egurtzegi et al., 2022), but crucially never as constituting separate peaks in looks to both the agent and patient referents. In Pitjantjatjara and Murrinhpatha, much less attention is paid to the agent in favor of a more even distribution of attention across the characters in the scene. The commonality between Pitjantjatjara and Murrinhpatha supports Nordlinger et al.'s suggestion that their results are attributable to the radically flexible word order in Murrinhpatha, rather than because it is polysynthetic.

We also found that relational encoding during event apprehension was influenced by word order. This was confirmed by our model comparison, where the model that included word order and its interaction with time was a better fit to the data. The differences across word orders are visualized in Fig. 5B, which suggests a distinction within event apprehension between A-initial and P-initial word orders from 400 ms onwards. Determining whether all four word orders differ from each other is a nontrivial process given the complexity of our models (e.g., by computing marginal effects), and so we have chosen to conservatively rely on the posterior probability plots for our interpretation (though see Appendix for posterior probability difference plots). It is not controversial to claim that word order influences sentence planning during event apprehension, since others have shown divergence in looks to the agent during this time window depending on word order (e.g., Norcliffe et al., 2015; Sauppe et al., 2013). Our analysis shows that, for Pitjantjatjara, word order influences relative looks to the agent versus the patient (and white space). What is unique about both Pitjantjara and Murrinhpatha is the pattern of early relational encoding that is affected by word order.

The data are consistent with the suggestion that sentence planning in Pitjantjatjara requires hierarchical planning, which is also consistent with our finding that there was no influence of participants' first fixation on word order selection. An important issue concerns the apparent difference in relational encoding between Australian languages and other languages with flexible word order, like Basque, Tseltal, and Tagalog. Our method was the same as that used by Norcliffe et al. (2015) for Tseltal and Sauppe et al. (2013) for Tagalog. While both languages allow multiple word order options, with Tagalog in particular allowing the free ordering of NPs in verb-initial sentences (though there are several conditioning factors, see Garcia & Kidd, 2022), both place some restrictions on verb placement such that they do not permit the type of flexibility available to speakers of Pitjantjatjara and Murrinhpatha. Thus, the difference in relational encoding in Pitjantjatjara and Murrinhpatha in comparison to Tseltal and Tagalog could reflect the greater need to quickly decide upon one word order from many possibilities, and reflect a difference in the degree of word order flexibility across these languages.

There are several lines of evidence in support of this suggestion. First, fixation patterns vary both within and across languages depending on the order of agent and patient arguments and in verb placement. Thus, in languages like English, Dutch, and German, there are comparatively more early fixations to patients than agents during the production of a passive sentence than an active sentence (Griffin & Bock, 2000; Konopka & Meyer, 2014; Sauppe, 2017). Similarly, in some circumstances, verb placement has been shown to affect fixation patterns early in sentence planning. In a study comparing English-speaking to Korean-speaking participants' eye-movements to causative scenes while they produced active transitive descriptions, Hwang and Kaiser (2014a) reported evidence of earlier verb encoding in English (as indexed by looks to an action region) than in Korean, which they interpreted to indicate that production in English is lexically driven, involving verb selection. In contrast, the fixations of Korean speakers were more evenly distributed across agent, patient, and action regions of the scene, with the results suggesting that verb lemma selection was delayed in comparison to English speakers, such that functional assignment of case can occur without verb selection (i.e., based on a conceptual representation of the event). Taken together, these results reveal how conceptualization differs depending on the linear sequencing of sentence constituents, demonstrating the flexible nature of sentence planning within and across languages (Ferreira & Swets, 2002). It is possible that, given the large number of word order options available to speakers of languages like Pitjantjatjara and Murrinhpatha, attention to event semantics and event roles (i.e., the entirety of the scene) smooths the transition into linguistic encoding.

A second source of evidence in favor of this point comes from the speech onset latencies. Although we did not analyze them, the speech latencies for Pitjantjatjara and Murrinhpatha were long in comparison to similar research. Notably, the average for both languages was always over 2300 ms for all word orders produced, which is more than 500 ms slower than average latencies in languages like German (Sauppe, 2017), Tagalog (Sauppe et al., 2013), Dutch, and Tseltal (Norcliffe et al., 2015). Not all of this difference can be attributed to the fact that Pitjantjatjara and Murrinhpatha speakers are unfamiliar with sentence production experiments, since Norcliffe et al.'s Tseltal speakers were similarly unfamiliar and showed similar speech onset latencies to Dutch speakers, although we cannot rule out the possibility that this contributes to some of the difference. Another logically possible variable is word order variability. There is evidence that syntactic flexibility delays speech onset across several languages (Russian: Myachykov, Scheepers, Garrod, Thompson, & Fedorova, 2013; Korean: Hwang & Kaiser, 2014b, for more evidence of competition in production, see Gennari, Mirkovic, & MacDonald, 2012; Haskell & MacDonald, 2003), a concept that is incorporated into most theoretical models of sentence production, where alternatives compete for selection based on processes like the functional properties of referents (e.g., their relative accessibility; their similarity to each other), language-specific structural constraints, overall structural frequency, and recent use (Chang, Dell, & Bock, 2006; Pickering & Branigan, 1998).

Our suggestion is that the lack of strong constraints on word order in languages like Pitjantjatjara and Murrinhpatha increases competition. In many circumstances, such as our unconstrained picture description task, where there is no clear discourse context to influence word order choice, this could result in an early need to decide upon a word order, driving the need to create an early conceptualization of the event to be described. However, given the large degree of competition between viable alternative word orders, linguistic encoding may still take some time, resulting in longer speech latencies. This may be particularly the case in experiments like ours, where there are no discourse cues for referent tracking.

How the large degree of word order flexibility characteristic of Australian languages influences speech onset times in naturalistic conversation is unclear. On the one hand, word order is significantly influenced by properties of discourse (Simpson & Mushin, 2008), and so discourse pressures will serve to promote some orders as more likely than others. On the other hand, it has often been observed that long inter-sentential silences are a distinctive feature of interaction among speakers of Australian Indigenous languages (e.g., Mushin & Gardner, 2009). It is difficult, however, to separate the observation of long pauses in conversation from cultural conventions for interaction, and so we leave the issue of how speaking a highly flexible word order language influences the chronometric properties of sentence formulation to further research.

If the free word order properties of Pitjantjatjara and Murrinhpatha drive early relational encoding, then one may question why Basque, which has been argued to have flexible or even free word order (Arantzeta et al., 2017; Erdocia, Laka, Mestres-Missé, & Rodriguez-Fornells, 2009), does not show a similar effect (Egurtzegi et al., 2022). We can identify two potential reasons, one linguistic and one methodological. Linguistically, while many case marking languages like Basque have flexible word orders, there is sufficient evidence in these languages for restrictions on ordering. Notably, there is clear evidence in Basque for a VP constituent following OV order (see Ros, Santesteban, Fukumura, & Laka, 2015; Santesteban et al., 2015). Australian languages are different in that there is very little evidence for such restrictions (Nordlinger, 2014); thus, the early relational encoding observed in Pitjantjatjara and Murrinhpatha may be a direct consequence of speaking such radically different languages.

Methodologically, it is possible that speakers of languages like Basque would show similar early relational encoding under more comparable experimental conditions, although the presence of constituency in flexible word order languages seems to place restrictions on sequencing that is not present in our data (see Myachykov, Thompson, Scheepers, & Garrod, 2011). We note that Egurtzegi et al.'s (2022) method constrained the ordering that their participants used, which may have limited the range of behavioral processes they could have observed. If they had allowed free description, they might have observed similar results to ours. At the same time, Egurtzegi et al. observed differences in the observed EEG power between Basque ergative transitive subjects and unmarked intransitive sentences, which they link to the early selection of verb lemmas, and which would require early relational encoding. We cannot compare those data to our eye-tracking data, in which verb encoding is often interpreted to be somewhat later and outside of event apprehension (Konopka, 2019). Given the different temporal resolutions of EEG and eye-tracking, and the inevitable ambiguity in linking both forms of measurement to language production processes, it may be that the methods capture early versus later processes underlying lexical selection for production, and thus our observation of early relational encoding in Pitjantjatjara, which we have interpreted to be mostly conceptual in nature, may indeed involve the early stages of linguistic encoding. If this is the case, then such grammatical encoding within event apprehension is not a unique feature of ergative or case marking languages, since our results align very closely with Nordlinger et al. (2022) Murrinhpatha data, which only infrequently uses case and prefers to mark participant roles via head marking in the polysynthetic verb (Nordlinger & Kidd, 2023). We are rather agnostic about this issue because we do not have good evidence either way, but note that the pattern of data we have observed is consistent with the conclusion that Pitjantjatjara speakers engage in a rapid form of relational encoding during event apprehension, a pattern which has only been observed for a similarly flexible word order language—Murrinhpatha.

4.3. *Limitations and future directions*

The current study has some limitations. First, it is only one study, although the close alignment between the current data and those of Nordlinger et al. (2022) for Murrinhpatha points to free word order having a significant effect on speech planning. Second, while the free production method we used has minimal task demands, its experimental nature reduced ecological validity enough to observe a distribution of word orders that differs from spontaneous Pitjantjatjara speech (Wilmoth et al., 2025). In response to this point, we note that this is a criticism that could be leveled at any elicited production study, and even though our word order distribution was different from that observed in spontaneous speech, the pattern of eye-movements during event apprehension still differed from past studies on other languages that have used the same or similar methods.

A final limitation concerns our analyses of the eye-tracking data. Following other studies, we did not include agent and patient humanness in our models analyzing eye-movements because to do so would have resulted in too few observations to make sensible conclusions, especially for the P-initial word orders. However, conceptual accessibility did have an influence on word order production, and so is likely to have influenced sentence planning.

5. Conclusion

Psycholinguistics has only haphazardly engaged with linguistic diversity (Collart, 2024; Berghoff & Bylund, 2025; Kidd & Garcia, 2022). This has led to an implicit assumption that theoretical concepts drawn from work on a handful of mostly European languages are universal. However, languages are incredibly diverse (Evans & Levinson, 2009), and we should expect to see the influence of this diversity in sentence production and processing mechanisms. In this paper, we have reported data to suggest that the ability to freely order sentence elements, a common feature of Australian Indigenous languages, significantly impacts sentence planning in a way that is independent of other typological differences across free word order languages. Specifically, our results suggest that free word order requires a rapid form of relational encoding that may differ from languages that place greater restrictions on word order. This is consistent with the idea that sentence planning in Pitjantjatjara is a hierarchical process, in which speakers rapidly appraise an event and quickly decide upon a word order to produce. The data add to a small but growing body of evidence that shows how prominent grammatical dimensions of languages shape the cognitive processes underlying language use.

Acknowledgments

This research was supported by grants from the Australian Research Council (CE140100041, DP210102836) and by the Max Planck Institute for Psycholinguistics. Additionally, this research was undertaken with the assistance of resources and services from the National Computational Infrastructure (NCI), which is supported by the Australian Government. We thank all members of the Pukatja community who helped and/or participated in the research, particularly Sandra Lewis, Agniezska Konopka, for sharing her materials, and Erin Walsh for drawing additional pictures. We thank Kiel Christianson and two anonymous reviewers for helpful comments, and Anton Malko for technical assistance.

Open access publishing facilitated by Australian National University, as part of the Wiley - Australian National University agreement via the Council of Australian University Librarians.

Open Research Badges

This article has earned Open Data and Open Materials badges. Data and materials are available at https://osf.io/xesfv/.

Notes

1 The "-Tu" represents a coronal stop with the same place of articulation as the preceding consonant, or /t/ following non-coronals.

2 ELAN (Version 6.4) [Computer software]. (2022). Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. Retrieved from https://archive.mpi.nl/tla/elan

3 In these cases, the eye-tracking software did not record the data. We attempted to find the source of the problem through multiple troubleshooting attempts but did not find the exact reason.

4 To compute track loss, we first calculated (for each participant and stimulus) the maximum duration of "no-looks" (within fixation runs per stimulus) at blank portions of the screen (i.e., the participant did not fixate any item or the eye-tracker did not record anything). For example, given a specific participant and stimulus, the duration between fixation runs where nothing was fixated was 200, 156, 456 ms, and so on. The maximum duration between fixation runs (for that participant in that specific stimulus) was 456 ms. Then, we created a sample with those values and assumed that the maximum duration between fixations follows a normal distribution. Thus, we took the observations between the sample mean +/− 3SDs, as this ensures that approximately 99% of the probability mass was taken into account. So, all the observations that did not fall within these values were removed.

5 For comparability with Nordlinger et al. (2022), we also analyzed the eye-movement data in the stages beyond event apprehension, using the same statistical models they used. These additional analyses are available on the study's OSF page.

References

Aissen, J. (1997). On the syntax of obviation. *Language*, *73*, 705–750. 10.1353/lan.1997.0042.

Arantzeta, M., Bastiaanse, R., Burchet, F., Wieling, M., Martinez-Zabaleta, M., & Laka, I. (2017). Eye-tracking the effect of word order in sentence comprehension in aphasia: Evidence from Basque, a free word order ergative language. *Language, Cognition, and Neuroscience*, *32*, 1230–1343. https://doi.org/10.1080/23273798.2017.1344715

Austin, P. K. (2001). Word Order in a Free Word Order Language: the case of Jiwarli. In J. Simpson and D. Nash, M. Laughren, P. Austin, & B. Alpher. (Eds.), *Forty Years On: Ken Hale and Australian Languages*, (pp. 305-324). Pacific Linguistics. doi: 10.15144/PL-512.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Barr, D. J. (2013). Random effects structure for testing interactions in linear mixed-effects models. *Frontiers in Psychology*, *4*, 328. https://doi.org/10.3389/fpsyg.2013.00328

Barr, D. J. (2008). Analyzing "visual world" eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, *59*, 457–474. https://doi.org/10.1016/j.jml.2007.09.002

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48. https://doi.org/10.18637/jss.v067.i01

Berghoff, R., & Bylund, E. (2025). Diversity in research on the psychology of language: A large-scale examination of sampling bias. *Cognition*, *256*, 106043. https://doi.org/10.1016/j.cognition.2024.106043

Blasi, D. E., Henrich, J., Adamou, E., Kemmerer, D., & Majid, A. (2022). Over-reliance on English hinders cognitive science. *Trends in Cognitive Sciences*, *26*, 1153–1170. https://doi.org/10.1016/j.tics.2022.09.015

Bock, K., & Levelt, W. J. M. (1994). Language production: Grammatical encoding. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 945–984). London: Academic Press.

Bock, K., Loebell, H., & Morey, R. (1992). From conceptual roles to structural relations: Bridging the syntactic cleft. *Psychological Review*, *99*, 150–171. https://doi.org/10.1037/0033-295X.99.1.150

Bock, K., Irwin, D. E., & Davidson, D. J. (2004). Putting first things first. In J. M. Henderson & F. Ferreira (Eds.), *The interface of language, vision, and action: Eye movements and the visual world* (pp. 249–278). Psychology Press.

Bock, J. K., & Warren, R.K. (1985). Conceptual accessibility and syntactic structure in sentence formulation. *Cognition*, *21*, (1), 47–67. https://doi.org/10.1016/0010-0277(85)90023-X

Bowe, H. J. (1990). *Categories, constituents, and constituent order in Pitjantjatjara: An aboriginal language of Australia*. London: Routledge.

Branigan, H. P., Pickering, M. J., & Tanaka, M. (2008). Contributions of animacy to grammaticla function assignment and word order during production. *Lingua*, *118*, 172-189. doi: 10.10116/j.lingua.2007.02.003

Brehm, L., & Alday, P. M. (2022). Contrast coding choices in a decade of mixed models. *Journal of Memory and Language*, *125*, 104334. https://doi.org/10.1016/j.jml.2022.104334

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using STAN. *Journal of Statistical Software*, *80*, (1), 1–28.

Bürkner, P.-C. (2018). Advanced Bayesian multilevel modeling with the R package brms. *R Journal*, *10*, (1), 395–411.

Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review*, *113*, (2), 234–272. https://doi.org/10.1037/0033-295X.113.2.234

Christianson, K., & Cho, H. Y. (2009). Interpreting null pronouns (pro) in isolated sentences. *Lingua*, *119*, (7), 989–1008. https://doi.org/10.1016/j.lingua.2008.11.005

Christianson, K., & Ferreira, F. (2005). Conceptual accessibility and sentence production in a free word order language (Odawa). *Cognition*, *98*, 105–135. https://doi.org/10.1016/j.cognition.2004.10.006

Collart, A. (2024). A decade of psycholinguistic research: Which place for linguistic diversity? *Glossa: Psycholinguistics*, *3*, 11. https://doi.org/10.5070/G60111432

Dixon, R. M. W. (1994). *Ergativity*. Cambridge: Cambridge University Press.

Egurtzegi, A., Blasi, D. E., Bornkessel-Schlesewsky, I., Laka, I., Meyer, M., Bickel, B., & Sauppe, S. (2022). Cross-linguistic differences in case marking shape neural power dynamics and gaze behavior during sentence planning. *Brain and Language*, *230*, 105127. https://doi.org/10.1016/j.bandl.2022.105127

Erdocia, K., Laka, I., Mestres-Missé, A., & Rodriguez-Fornells, A. (2009). Syntactic complexity and ambiguity resolution in a free word order language: Behavioral and electrophysiological evidences from Basque. *Brain and Language*, *109*, 1–17. https://doi.org/10.1016/j.bandl.2008.12.003

Esaulova, Y., Penke, M., & Dolscheid, S. (2019). Describing events: Changes in eye movements and language production due to visual and conceptual properties of scenes. *Frontiers in Psychology*, 10. Article 835. https://doi.org/10.3389/fpsyg.2019.00835

Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences*, *32*, 429–492. https://doi.org/10.1017/S0140525×0999094X

Fahrmeir, L., & Kneib, T. (2011). *Bayesian smoothing and regression for longitudinal, spatial and event history data*. Oxford University Press.

Ferreira, F., & Swets, B. (2002). How incremental is language production? Evidence from the production of utterances requiring the computation of arithmetic sums. *Journal of Memory and Language*, *46*, 57–84. https://doi.org/10.1006/jmla.2001.2797

Ferreira, V. S., & Slevc, L. R. (2007). Grammatical encoding. In G. Gaskell (Ed.), *Oxford handbook of psycholinguistics* (pp. 453–470). Oxford: Oxford University Press.

Garret, M. F. (1975). The analysis of sentence production. *Psychology of Learning and Motivation*, *9*, 133–177. https://doi.org/10.1016/S0079-7421(08)60270-4

Garcia, R., & Kidd, E. (2022). Acquiring verb-argument structure in Tagalog: A multivariate corpus analysis of caregiver and child speech. *Linguistics*, *60*(6), 1855-1906. 10.1515/ling-2021-0107.

Garcia, R., Roeser, J., & Kidd, E. (2023). Finding your voice: Voice-specific effects in Tagalog reveal the limits of word order priming. *Cognition*, 236. Article 105424. https://doi.org/10.1016/j.cognition.2023.105424

Gennari, S. P., Mirkovic, J., & MacDonald, M. C. (2012). Animacy and competition in relative clause production: A cross-linguistic investigation. *Cognitive Psychology*, *65*, 141–176. https://doi.org/10.1016/j.cogpsych.2012.03.002

Gerwien, J., & Flecken, M. (2016). First things first? Top-down influences on event apprehension. In A. Papafragou, D. Grodner, D. Mirman, & J. Trueswell (Eds.), *Proceedings of the 38th Annual Meeting of the Cognitive Science Society* (CogSci 2016) (pp. 2633–2638). Austin, TX: Cognitive Science Society.

Gleitman, L. R., January, D., Nappa, R., & Trueswell, J. C. (2007). On the give and take between event apprehension and utterance formulation. *Journal of Memory and Language*, *57*, 544–569. https://doi.org/10.1016/j.jml.2007.01.007

Goddard, C. (1982). Case systems and case marking in Australian languages: A new interpretation. *Australian Journal of Linguistics*, *2*, 167–196. https://doi.org/10.1080/07268608208599290

Goddard, C. (1988). Verb serialisation and the circumstantial construction in Yankunytjatjara. In P. Austin (Ed.), *Complex sentence constructions in Australian languages* (pp. 177–192). Amsterdam: Benjamins.

Goddard, C., & Harkins, J. (2002). Posture, location, existence, and states of being in two Central Australian languages. In J. Newman (Ed.), *The linguistics of sitting, standing, and lying* (pp. 213–238). Ebsco Publishing. Retrieved from http://site.ebrary.com/id/10046377

Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science*, *11*, 274–279. https://doi.org/10.1111/1467-9280.00255

Hale, K. (1983). Warlpiri and the grammar of nonconfigurational languages. *Natural Language and Linguistic Theory*, *1*(1), 5–47. https://doi.org/10.1007/BF00210374

Haskell, T. R., & MacDonald, M. C. (2003). Conflicting cues and competition in subject-verb agreement. *Journal of Memory and Language*, *48*, 760–778. https://doi.org/10.1016/S0749-596X(03)00010-X

Hastie, T., & Tibshirani, R. (1990). *Generalized additive models*. London: Chapman & Hall.

Henrich, J., Heins, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral & Brain Sciences*, *33*, 61–83. https://doi.org/10.1017/S0140525×0999152X

Himmelmann, N. P. (2005). Tagalog. In A. Adelaar & N. P. Himmelmann (Eds.), *The Austronesian languages of Asia and Madagascar* (pp. 350–376). Routledge.

Hwang, H., & Kaiser, E. (2014a). The role of the verb in grammatical function assignment in English and Korean. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*, 1363–1376. https://doi.org/10.1037/a0036797

Hwang, H., & Kaiser, E. (2014b). Having syntactic choice is not always better: The effects of syntactic flexibility on Korean production. *Language, Cognition, and Neuroscience*, *29*, 1115–1131. https://doi.org/10.1080/23273798.2013.875212

Hwang, H., & Kaiser, E. (2015). Accessibility effects on production vary cross-linguistically: Evidence from English and Korean. *Journal of Memory and Language*, *84*, 190–204. https://doi.org/10.1016/j.jml.2015.06.004

Isasi-Isasmendi, A., Andrews, C., Flecken, M., Laka, I., Daum, M. M., Meyer, M., Bickel, B., and Sauppe, S. (2023). The agent preference in visual event apprehension. *Open Mind*, *7*:240–282. doi: 10.1162/opmi_a_00083

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*, 434–446. https://doi.org/10.1016/j.jml.2007.11.007

Jaeger, T. F., & Norcliffe, E. J. (2009). The cross-linguistic study of sentence production. *Language and Linguistics Compass*, *3*, 866–887. https://doi.org/10.1111/j.1749-818X.2009.00147.x

James, G., Witten, D., Hastie, T., Tibshirani, R., & Taylor, J. (2023). *An introduction to statistical learning: With applications in Python* (1st ed.). Cham Springer International Publishing.

Kidd, E., & Garcia, R. (2022). How diverse is child language acquisition research? *First Language*, *42*, 703–735. https://doi.org/10.1177/01427237211066405

Konopka, A. E., & Meyer, A. S. (2014). Priming sentence planning. *Cognitive Psychology*, *73*, 1-40. 10.1016/j.cogpsych.2014.04.001.

Konopka, A. E. (2019). Encoding actions and verbs: Tracking the time-course of relational encoding during message and sentence formulation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(8), 1486–1510. https://doi.org/10.1037/xlm0000650

Konopka, A. E., & Brown-Schmidt, S. (2014). Message encoding. In V. Ferreira, M. Goldrick, & M. Miozzo (Eds.), *The Oxford handbook of language production* (pp. 3–20). New York: Oxford University Press.

Konopka, A., Meyer, A. S., & Forest, T. A. (2018). Planning to speak in L1 and L2. *Cognitive Psychology*, *102*, 72–104. https://doi.org/10.1016/j.cogpsych.2017.12.003

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*, 1–38. https://doi.org/10.1017/S0140525×99001776

Marmion, D., Obata, K., & Troy, J. (2014). *Community, identity, wellbeing: The report of the Second National Indigenous Languages Survey*. Canberra: Australian Institute of Aboriginal and Torres Strait Islander Studies.

Mujkic, E. (2013). *Murrinh-Patha syntax: Clausal structure and the noun phrase*. University of Melbourne Honours thesis.

Mushin, I., & Gardner, R. (2009). Silence is talk: Conversational silence in Australian talk-in-interaction. *Journal of Pragmatics*, *41*, 2033–2052. https://doi.org/10.1016/j.pragma.2008.11.004

Myachykov, A., Garrod, S., & Scheepers, C. (2010). Perceptual priming of structural choice during English and Finnish sentence production. In R. K. Mishra & N. Srinivasan (Eds.), *Language and cognition: State of the art* (pp. 54–72). Munich: Lincom Europa.

Myachykov, A., Scheepers, C., Garrod, S., Thompson, D., & Fedorova, O. (2013). Syntactic flexibility and competition in sentence production: The case of English and Russian. *Quarterly Journal of Experimental Psychology*, *66*, 1601–1619. 10.1080/17470218.2012.754910

Myachykov, A., Thompson, D., Scheepers, C., & Garrod, S. (2011). Visual attention and structural choice in sentence production across languages. *Language and Linguistic Compass*, *5*, 95–107. https://doi.org/10.1111/j.1749-818X.2010.00265.x

Norcliffe, E., & Jaeger, T. F. (2016). Predicting head-marking variability in Yucatec Maya relative clause production. *Language and Cognition*, *8*, 167–205. https://doi.org/10.1017/langcog.2014.39

Norcliffe, E., Konopka, A. E., Brown, P., & Levinson, S. C. (2015). Word order affects the time course of sentence formulation in Tzeltal. *Language, Cognition and Neuroscience*, *30*, 1187–1208. https://doi.org/10.1080/23273798.2015.1006238

Nordlinger, R. (2014). Constituency and grammatical relations in Australian languages. In H. Koch & R. Nordlinger (Eds.), *The language and linguistics of Australia: A comprehensive guide* (pp. 215–261). Berlin: Walter de Gruyter.

Nordlinger, R., & Bresnan, J. (2011). Lexical-functional grammar: Interactions between morphology and syntax. In R. D. Borsley & K. Börjars (Eds.), *Non-transformational syntax: Formal and explicit models of grammar* (pp. 112–140). Oxford: Wiley-Blackwell.

Nordlinger, R., Garrido Rodriguez, G., & Kidd, E. (2022). Sentence planning and production in Murrinhpatha, an Australian 'free word order' language. *Language*, *98*, 187–220. Retrieved from https://muse.jhu.edu/article/857152

Nordlinger, R., & Kidd, E. (2023). Uncovering ergative use in Murrinhpatha: Evidence from experimental data. *Australian Journal of Linguistics*, *43*, 69-83. https://doi.org/10.1080/07268602.2023.2222086

Pickering, M. J., & Branigan, H. P. (1998). The representation of verbs: Evidence from syntactic priming in language production. *Journal of Memory and Language*, *39*(4), 633–651. https://doi.org/10.1006/jmla.1998.2592

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS*. Spinger.

Porretta, V., Kyröläinen, A., van Rij, J., & Järvikivi, J. (2018). Visual world paradigm data: From preprocessing to nonlinear time-course analysis. In I. Czarnowski, R. J. Howlett, & L. C. Jain (Eds.), *Intelligent design technologies 2017: Proceedings of the 9th KES International Conference on Intelligent Decision Technologies (KES-IDT 2017)* (pp. 268–277). New York: Springer. https://doi.org/10.1007/978-3-319-59424-8_25

R Core Team (2024). R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing*, Vienna, Austria. https://www.R-project.org/.

Roelofs, A., & Ferreira, V. S. (2019). The architecture of speaking. In P. Hagoort (Ed.), *Human language: From genes and brains to behavior* (pp. 35–50). Cambridge, MA: MIT Press.

Ros, I., Santesteban, M., Fukumura, K., & Laka, I. (2015). Aiming at shorter dependencies: The role of agreement morphology. *Language, Cognition and Neuroscience*, *30*, 1156–1174. doi:10.1080/23273798.2014.994009

Santesteban, M., Pickering, M. J., Laka, I., & Branigan, H. P. (2015). Effects of case-marking and head position on language production. Evidence from an ergative OV language. *Language, Cognition, and Neuroscience*, *30*, 1175–1186. 10.1080/23273798.2015.1065335

Sassenhagen, J., & Alday, P. M. (2016). A common misapplication of statistical inference: Nuisance control with null-hypothesis significance tests. *Brain & Language*, *162*, 42–45.

Sauppe, S. (2017). Word order and voice influence the timing of verb planning in German sentence production. *Frontiers in Psychology*, *8*, 1648. https://doi.org/10.3389/fpsyg.2017.01648

Sauppe, S., Norcliffe, E. J., Konopka, A. E., Van Valin, R. D., & Levinson, S. C. (2013). Dependencies first: Eye-tracking evidence from sentence production in Tagalog. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 265–1270). Austin, TX: Cognitive Science Society.

Sauppe, S., Choudhary, K. K., Giroud, N., Blasi, D. E., Norcliffe, E., Bhattamishra, S., Gulati, M., Egurtzegi, A. M., Bornkessel-Schlesewsky, I., Meyer, M., and Bickel, B. (2021). Neural signatures of syntactic variation in speech planning. *PLOS Biology*, *19*(1):e3001038. doi: 10.1371/journal.pbio.3001038

Schad, D. J., Vasishth, S., Hohenstein, S., & Kliegl, R. (2020). How to capitalize on contrasts in linear (mixed) models: A tutorial. *Journal of Memory and Language*, *110*, 104038. https://doi.org/10.1016/j.jml.2019.104038

Schlenter, J., Esaulova, Y., Dolscheid, S., & Penke, M. (2022). Case ambiguity does not affect the description of events in German: Evidence from sentence production and eye-tracking. *Language, Cognition, and Neuroscience*, *37*, 844 - 865. https://doi.org/10.1080/23273798.2022.2026419

Skirgård, H., Haynie, H. J., Blasi, D. E., Hammarström, H., Collins, J., Latarche, J. J., … Gray, R. D. (2023). Grambank reveals the importance of genealogical constraints on linguistic diversity and highlights the impact of language loss. *Science Advances*, *9*(16), eadg6175. https://doi.org/10.1126/sciadv.adg6175

Silverstein, M. (1976). Hierarchy of features and ergativity. In R. M. W. Dixon (Ed.), *Grammatical categories in Australian languages* (pp. 112–171). Canberra: Australian Institute for Aboriginal Studies.

Simpson, J., & Mushin, I. (2008). Clause-initial position in four Australian languages. In I. Mushin & B. Baker (Eds.), *Discourse and grammar in Australian languages* (pp. 25–57). Amsterdam: John Benjamins.

Tanaka, M. N., Branigan, H. P., McLean, J. F., & Pickering, M. J. (2011). Conceptual influences on word order and voice in sentence production: Evidence from Japanese. *Journal of Memory and Language*, *65*, 318–330. https://doi.org/10.1016/j.jml.2011.04.009

Vaughan, J., & Singer, R. (2018). Indigenous multilingualisms past and present. *Language & Communication*, *62*, 83–90. https://doi.org/10.1016/j.langcom.2018.06.003

Wilmoth, S. (2023). *The dynamics of contemporary Pitjantjatjara: An intergenerational study*. PhD Thesis. University of Melbourne.

Wilmoth, S., Garrido Rodriguez, G., Nordlinger, R., & Kidd, E. (2025). Word order flexibility in Pitjanjatjara across genres and generations. In C. O'Shannassy, J. Bednall, & D. Angelo (Eds.), *Projecting voices: Studies in language and linguistics in honour of Jane Simpson*. Canberra: ANU Press.

Wood, S. N. (2017). *Generalized additive models: An introduction with R* (2nd ed.). London: CRC Press.

Zuur, A., Ieno, E. N., Walker, N., Saveliev, A. A., & Smith, G. M. (2009). *Mixed effects models and extensions in ecology with R*. New York: Springer. https://doi.org/10.1007/978-0-387-87458-6.

Appendix

Fig. A1 shows the posterior probability difference plots that depict pairwise comparisons in fixations to the agent across word orders during event apprehension. We calculated the predicted probabilities of fixating the agent for each word order based on the following criteria: (1) Participant 1 was used as the reference, (2) agent-size and patient-size were held at their respective means, and (3) the analysis focused on a specific stimulus = army_general_kicking_boyL_T.jpg.

We then computed the difference of fixating the agent between all word order pairs (i.e., the difference between fixating the agent in APV vs. AVP, APV vs. PAV, etc.). The reference word order level appears on the left side of the comparison (i.e., either APV or PAV). Ribbons indicate pointwise 95% credible intervals. If the line and ribbon are higher than 0, there are more looks to the agent in the reference-level word order. Thus, for the APV-AVP comparison, there are more looks to the agent around 350 ms in APV compared to AVP, and so on. We see that there are differences for all comparisons except for the two P-initial word orders, where looks to the agent are not predicted to differ. These plots reinforce our conclusion that word order significantly influences sentence planning in event apprehension in Pitjantjatjara.

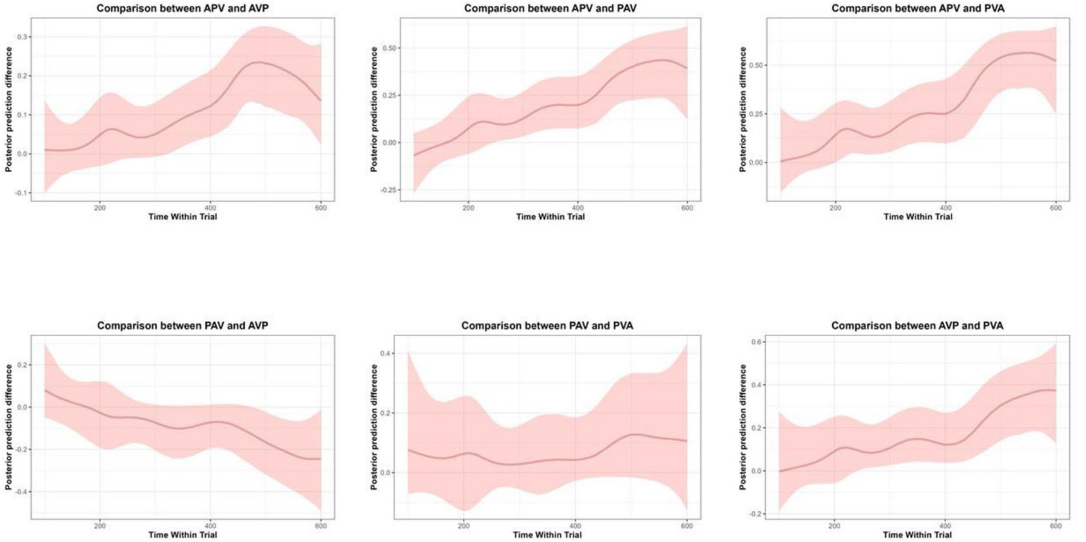

Fig. A1. Posterior probability difference plots of fixating the agent across different word order comparisons over time. Ribbons indicate pointwise 95% credible intervals.