16

Word order flexibility in Pitjantjatjara

Sasha Wilmoth, Rachel Nordlinger, Gabriela Garrido Rodríguez and Evan Kidd

Introduction

Word order flexibility as a feature of (some) Australian languages has been much discussed in the theoretical literature since Hale's early work on Warlpiri (e.g. Hale 1981, 1983; see also Simpson 1991; Austin and Bresnan 1996; Nordlinger and Bresnan 2011), and it has long been recognised that word order does not play a major role in these languages in identifying grammatical functions (Hale 1983; Nordlinger 1998; Simpson 1983). Simpson (1983, 18) describes Warlpiri as a language in which 'the burden of representing the relations between predicates and arguments … is borne by the morphology rather than the syntax', highlighting the importance of case marking and argument agreement for encoding grammatical relations, as opposed to word order. Instead, word order in these languages is said to be 'pragmatically' driven, determined by information packaging and discourse prominence (Swartz 1991; Rose 2001; Simpson 2007; Simpson and Mushin 2008). However, despite the longstanding interest in their flexible syntax, there have been surprisingly few detailed studies of word order flexibility in Australian languages (cf. Swartz 1991; Austin 2001), and fewer still that take into consideration variation between genres or between generations of speakers.

In this chapter, we seek to address this gap with a discussion of word order in Pitjantjatjara, drawing on both naturalistic and experimental data recorded by the first author in Pukatja/Ernabella in 2018–19, from 67 L1 speakers. The use of experimental data in addition to a more naturalistic corpus allows us to investigate different syntactic or semantic drivers of word order, which can be confounded by the information packaging requirements of naturalistic discourse. Pitjantjatjara has been described as having a basic order of SOV but with much flexibility and frequent omission of arguments (Bowe 1990). In our study, we are interested in the range of possible word orders found across the different genres and what conditions appear to determine the choice of one word order over another. We show that word order choice in Pitjantjatjara is neither solely syntactic/semantic (based on semantic role and/or syntactic function), nor solely pragmatic; rather, it is determined by a complex array of interacting factors, including grammatical function, the humanness of event characters and speaker preference. We also consider word order flexibility across generations. Following Langlois (2004), who claimed young Pitjantjatjara speakers at Utju/Areyonga in the 1990s were undergoing a shift from basic SOV to SVO order as a result of contact with English, we investigated age-related differences in word order use. Interestingly, our data suggest that rates of word order flexibility and use of SOV as the most common word order are stable across the generations, with no clear evidence that younger generations are doing anything different to their elders in these respects. While the naturalistic corpus only includes female speakers, there were both male and female participants in the experimental study, allowing us to discuss some gender differences. This research adds to the small literature on word order variability in Australian languages and advances our understanding of how word order is determined in languages with very flexible word order. Our discussion of differences between generations also adds to our understanding of language variation and change in Australian languages post-colonisation.

Overview of Pitjantjatjara

Pitjantjatjara is a Western Desert (Pama-Nyungan) variety spoken in Central Australia. It is the dominant language across the Aṉangu Pitjantjatjara Yankunytjatjara (APY) Lands in far north South Australia and some South Australian communities outside this region, such as Yalata. It is also spoken in some communities in the Northern Territory and Western Australia, such as Irrunytju, Muṯitjulu and Utju (Areyonga). Pitjantjatjara is learned as a first language by most children in these communities and in the APY Lands, and it is spoken between Aṉangu (Aboriginal people) in most domains of everyday life. As of the 2016 census, there are over 3,000 people who speak Pitjantjatjara at home (DITRDC, AIATSIS and ANU 2020). A map of the tristate region is shown in Figure 16.1. This map includes communities in Western Australia who speak other Western Desert varieties such as Ngaanyatjarra and Ngaatjatjarra; Yankunytjatjara is also an important variety spoken in the APY Lands, although it is more endangered than Pitjantjatjara (Naessan 2008).

Pitjantjatjara is a dependent-marking language, with obligatory ergative marking on nominals except for pronouns, which follow a nominative-accusative pattern. Grammatical relations are determined by case marking. There is no marking of participants on verbs; arguments may also be indexed by second position clitic pronouns, although these are not obligatory and not all person/number/case combinations have clitic pronoun equivalents available.

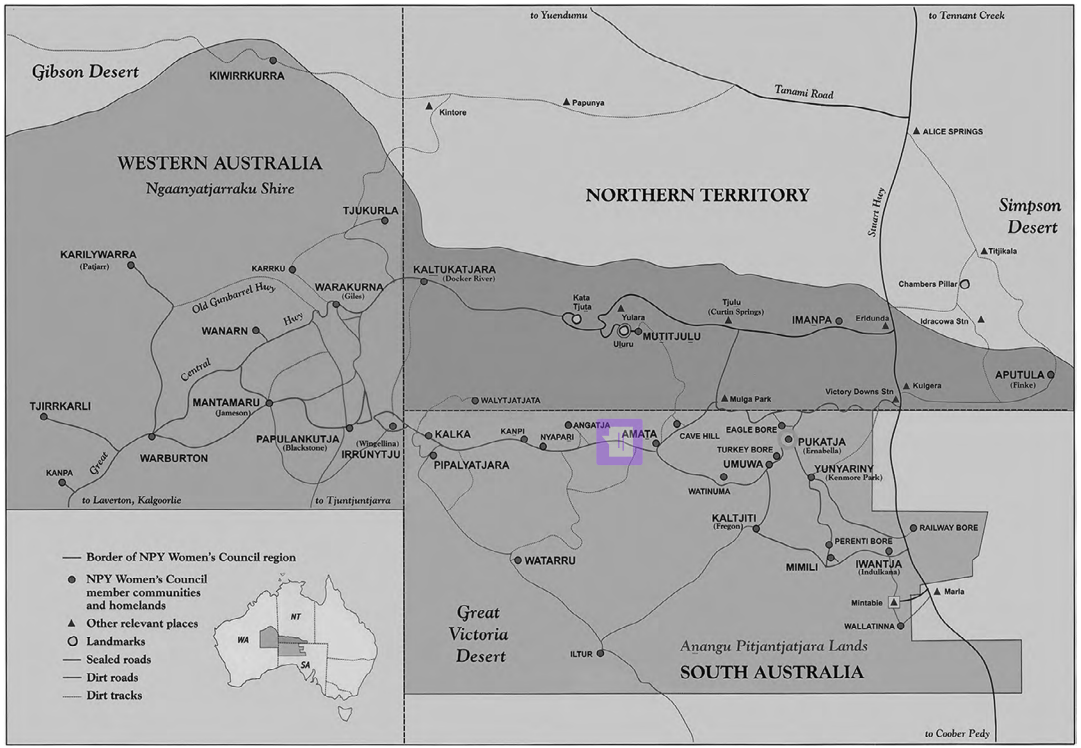

Figure 16.1: Map of Ngaanyatjarra Pitjantjatjara Yankunytjatjara (NPY) region.

Note: The location of Pukatja is circled in white.

Source: Courtesy of NPY Women's Council (www.npywc.org.au/about/).

Previous research on word order and ellipsis in Pitjantjatjara

The main sources on word order and ellipsis in Pitjantjatjara, which we summarise here, are Bowe (1990), Langlois (2004) and Goddard (1985), who all discuss word order and ellipsis as part of broader descriptions of the language. A notable absence in the literature, which is outside the scope of this chapter, is a comprehensive account of prosody and constituent order (although see Tabain and Fletcher 2012 for a preliminary analysis of intonation, and Defina et al. 2020 for an analysis of prosody in clause chains). Pitjantjatjara has been described as having a basic order of Subject-Object-Verb (SOV), but with much flexibility and frequent omission of arguments (Bowe 1990). Langlois (2004) describes a change to a basic order of SVO due to contact with English among teenage speakers of Pitjantjatjara in Areyonga (a community some 400 kilometres to the north of Pukatja, where all the data we discuss were recorded).[1] Both Bowe and Langlois counted the frequency of different word orders in relatively small samples of narrative data. Bowe's sample includes 13 (presumably monologic) narratives from adult women and one adult man; these stories include five historical narratives, seven children's stories and one Dreaming excerpt (Bowe, 1990, 116). Langlois's sample includes 10 stories from teenage girls between 14 and 18 years old. Their results regarding transitive clauses are shown in Table 16.1. However, Langlois's count only includes transitive clauses with at least one 'full' (non-pronominal) NP (2004, 80), so these two sets of figures are not entirely comparable.

Table 16.1 shows that both researchers found a range of word orders across their corpus, although Langlois's dataset is fairly small, so it is less likely to capture less frequently occurring orders. The most frequent orders in Bowe's (1990) dataset are V, OV and SV, with a strong overall preference for NP omission. Bowe (1990) argues that SOV is the basic order for several reasons. It was the most frequent order among clauses where both subject and object NPs are expressed, with more than twice as many occurrences as SVO. Bowe also notes that some consultants 'criticised' sentences with SVO order during transcription, saying that the speaker 'forgot' to say the object pre-verbally (1990, 118).

1 Bowe's fieldwork was in Mimili (South Australia) in the 1980s, while Langlois's fieldwork was undertaken in 1994–95 (although published in 2004). Goddard's work on Yankunytjatjara, discussed below, was conducted in Mimili as well as other communities in the APY Lands.

Table 16.1: Word orders in transitive sentences in narrative data recorded by Bowe (1990) with adult speakers in Mimili and Langlois (2004) with teenage speakers in Areyonga.

	Bowe (1990)		Langlois (2004)	
SOV	17	5%	16	27%
SVO	8	2%	24	41%
SV	58	17%	11	19%
OSV	3	1%	–	–
OVS	5	2%	–	–
OV	57	17%	7	12%
VSO	–	0%	–	–
VOS	–	0%	–	–
VS	4	1%	–	–
VO	6	2%	1	2%
V	173	52%	–	–
Total	333	100%	59	100%

As well as postulating SOV as the basic order, Bowe discusses some syntactic and pragmatic features of OSV, SVO and OVS sentences. She argues that OSV order 'may result from a topicalisation of the object', in which case it is usually 'crucially linked' to the previous clause (1990, 132–33). However, the nature of this link is not defined. SVO sentences can arise where the object is an 'afterthought or anti-topic' (p. 131), in which case it is preceded by a pause and is 'syntactically extra-sentential'. When there is no pause, and the object is prosodically part of the rest of the clause, Bowe suggests it may be a verb-prominent construction. This occurs, for example, where the object is clearly established in the discourse, and it is the action itself that is the most important piece of information (p. 132). OVS occurs where the subject is 'an afterthought or anti-topic' (p. 133), usually following a pause.

Further discussion of factors bearing upon word order variability in the language is in Goddard's (1985) grammar of Yankunytjatjara (a dialect very closely related to Pitjantjatjara). Although he does not include any quantitative analysis of corpus data, he notes that non-SOV orders arise because the 'sentence-initial position reflects "prominence" or speakers' focus of interest' (p. 21). Like Bowe, he argues that post-verbal NPs are afterthoughts, often appearing following a pause, or that the information is 'peripheral to the speaker's main concerns' (p. 21).

Related to issues of word order variability is ellipsis of NP arguments, which is ubiquitous in Pitjantjatjara discourse, as seen in Bowe's results above. Goddard notes that third person singular NPs can be freely omitted if their reference is clear from context, while first and second person NPs are less likely to be omitted (1985, 22) and appear as clitic pronouns in such contexts (there is no equivalent clitic pronoun for third person singular subjects or objects). Bowe (1990) argues that omission of objects is more 'grammatically marked' and less preferred by speakers than omission of subjects, potentially due to switch reference distinctions in the grammar making elided subjects more recoverable than elided objects (p. 66–67). Goddard also makes this connection between ellipsis of subjects and switch reference; however, he notes that 'most ellipsis affects objects' (1985, 21).

Langlois (2004) does not offer any discussion of particular factors driving word order variability in her data, although she argues that the predominance of SVO as opposed to the more traditional SOV is due to increased contact with English (p. 82). Her data, while suggestive of an ongoing change to basic SVO order, are not based on a very large sample and only include teenage speakers; more research is needed to understand if this is a robust tendency. Despite this apparent shift to a preference for SVO, there is no change in how grammatical relations are signalled in Pitjantjatjara in her data. This is not the case for several other Australian Aboriginal languages in language shift situations, many of which have lost traditional systems of argument marking (such as ergative case marking or bound pronouns on verbs) as word order has rigidified to SVO (Bavin and Shopen 1985; Lee 1987; McGregor 2002; Meakins and O'Shannessy 2010; Pensalfini 1999; Richards 2001; Schmidt 1985).

These previous studies of word order in Pitjantjatjara point to a large amount of variability, certainly in choice of word order, but also choice or absence of referring expression (ellipsis, clitic pronoun, free pronoun, or lexical NP, with or without the definite third person singular pronoun). This variability is related to a complex mix of factors, not all of which are discussed above. These include prosody, discourse prominence, topic/focus, (re)introduction of referents, whether it is a main or subordinate clause, generation of speakers, genre, language contact, switch reference and reported speech. However, until now, there have been no further detailed studies of word order at a larger scale in any variety of Pitjantjatjara or Yankunytjatjara. Out of this huge range of variables and possibilities, in this chapter, we specifically focus on word order variability, including ellipsis, with respect to subject, object and verb in transitive, finite main

clauses. Given the limitations of previous studies, here we try to describe and quantify word order variability based on a larger dataset, and using a combination of experimental and more naturalistic data.

Methods and data collection

We collected two types of data: speakers' responses in an experimental picture-description task, and spontaneous speech (primarily narratives of different types). Each type of data has its strengths and weaknesses in terms of understanding the drivers behind word order variability. Experimental data allow us to manipulate variables that likely condition variation, in our case, the conceptual accessibility of agent and patient referents, which typically follow skewed distributions in spontaneous data, making them more difficult to analyse. It also controls other factors that are also likely to have an effect, such as discourse structure, which is naturally highly variable in spontaneous speech. Thus, a comparison of the two types of data allows us to understand the relative contribution of these different factors. The data were collected in Pukatja, also known as Ernabella, in the Aṉangu Pitjantjatjara Yankunytjatjara (APY) Lands in South Australia in 2018–19. The majority of speakers were residents in Pukatja at the time, although a handful were visiting from other nearby Pitjantjatjara-speaking communities such as Muṯitjulu and Amaṯa.

Participants provided informed consent before recording (information was provided both orally and in writing, in a combination of Pitjantjatjara and English, depending on the individual participant). The projects were approved by the University of Melbourne Human Research Ethics Council (projects 1647234 and 1953646.1). Participants were paid for their time. All data was transcribed in ELAN[2] by the first author with the assistance of a Pitjantjatjara-speaking collaborator, where needed. Transcription includes variation in pronunciation and morphology; thus, some words are not spelled in the same form as they appear in the dictionary and other sources. Details specific to each type of data are provided below.

2 *ELAN*, v6.3, Max Planck Institute for Psycholinguistics, The Language Archive, 2022, PC, archive.mpi.nl/tla/elans.

Experimental data

A picture description task (see Norcliffe et al. 2015; Nordlinger, Garrido Rodriguez and Kidd 2022) was used to elicit participants' descriptions of different transitive events. Participants were shown 48 experimental images; an example is shown in Figure 16.2. The test pictures fully crossed thematic role (agent, patient) with humanness (human, non-human), such that 12 pictures had a human agent and a human patient (*girl pushes boy*), 12 had a human agent and non-human patient (*man catches fish*), 12 had a non-human agent with a human patient (*crocodile chases people*) and 12 paired a non-human agent with a non-human patient (*dog chases car*). There were two mirror-reversed versions of each experimental item: one in which the agent appeared on the left-hand side of the visual display and another in which it appeared on the right-hand side of the picture. An additional 96 pictures served as fillers between the test pictures, mostly depicting single-argument events. Four blocks of stimuli were created, each containing 36 items: 12 test and 24 fillers. Participants were tested individually and were given the opportunity to take a short break in between blocks. The blocks were shown in different orders to different participants, to balance any potential effects of seeing an image at the beginning or the end of the task. The visual displays were presented in a Lenovo ThinkPad laptop computer (13″ monitor), and participants' picture descriptions were recorded using a Zoom H1 Handy recorder.

The researcher was sitting perpendicular to the participant in order to control the computer; both the participant and the researcher were able to view the images on the computer screen. Thus, the pragmatic context of the experiment was such that the speaker was describing the picture in the presence of an audience that could already see it. All images depicted unique event characters, and there was no relationship between the different pictures that would suggest a narrative or discourse context. Thus, in this sense, all referents were 'new' in that they had not been mentioned in any preceding discourse; however, they were accessible to both speaker and addressee. It is difficult to know what, if any, impact this may have had on participant responses, but the key point for our purposes is that the pragmatic conditions were held constant across all trials; thus, variation in word order cannot be attributed to variation in information status or other discourse-pragmatic factors. Eye movements were also recorded; these are not the focus of this chapter, and the details for this are described in Kidd et al. (2025). Forty-nine native speakers of Pitjantjatjara (30 females, *M* age = 36.81, *SD* = 11.98, range 18–60 years old) participated in the experimental study.

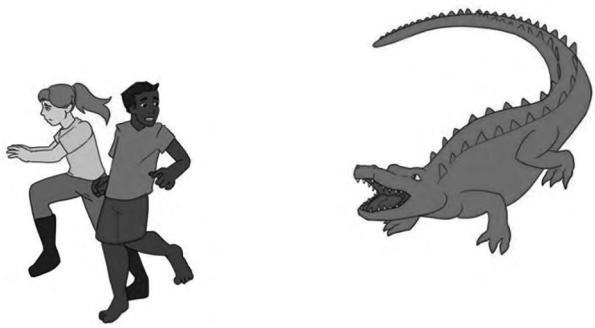

Figure 16.2: One of the images used in the experimental picture description task.
Source: Authors.

Of the potential 2,352 observations (49 participants × 48 stimuli images), the initial dataset consisted of 2,339 observations (two participants did not complete the full experiment). From this pool of 2,339 observations, responses were excluded if participants did not describe the event with a simple transitive sentence or if they self-corrected (in which case, the entire utterance was excluded). Examples of the types of exclusions are shown in (1a–c).

(1)a. Two clauses, one intransitive to introduce the agent:

tjati	*kutju*	*ngari-nyi*	*ka*	*tjitji*	*kutju-ngku*	*ngaṟa-la*
lizard	one	lie-PRS	and.DS[3]	child	one-ERG	stand-MV

pu-nga-nyi
hit-AUG-PRS
'A lizard is lying and a child is standing and hitting [it].' (20181007-PJ32)

b. Self-correction:

papa	*riitja-ma-ṉu*	…	*papa-ngku*	*riitja-ma-ṉu*	*rapita*
dog	run-LOAN-PST	…	dog-ERG	run-LOAN-PST	rabbit

'The dog[NOM/ACC] chased … the dog[ERG] chased the rabbit.' (20181002-PJ19)

3 Abbreviations not in Leizpg Glossing Rules: AUG – verbal augment; DS – different subject; INTENT – intentive; LOAN – transitive loanword verbaliser; MV – medial verb; SS – same subject

c. Intransitive clause with intended patient not encoded as an accusative object:

papa *inka-nyi* *butterfly-ngka*
dog play-PRS butterfly-LOC
'The dog is playing with the butterfly.' (20180925-PJ44)

This left 1,865 transitive clauses. Each of these was coded for speaker, speaker age and gender, word order and humanness of agent and patient. All agents/patients were subjects/objects, and all subjects/objects were agents/patients.

Spontaneous data

The spontaneous data set consisted of 49 recordings made in Pukatja in 2019, as part of the first author's PhD project that investigated continuity and change between generations of Pitjantjatjara speakers (Wilmoth 2022). Spontaneous speech was recorded from 22 female speakers (*M* age = 34.28, *SD* = 16.62, range 17–69 years old). The spontaneous data consisted of a combination of narratives (monologic or co-constructed), elicited narratives (from wordless storybooks or videos),[4] or the 'Family Problem' elicitation task (San Roque et al. 2012). From 2,821 utterances, there were 1,215 transitive clauses that were selected initially for analysis. Of these, 105 were excluded because they were transitive clauses in a clause chain where the finite verb was intransitive, or imperative or interrogative utterances (as these cannot be compared to the experimental data which only contained declarative utterances), or clauses where a core argument was expressed as a pronominal clitic (as these are always in a fixed position so don't contribute to a study of word order flexibility). The total number of observations for analysis was thus 1,110 transitive clauses. These were coded for speaker, speaker age and gender, word order and agent and patient humanness. Several of the narrative elicitation materials included anthropomorphised animals as characters (such as the penguin characters in the animated show *Pingu*). These patterned in a similar way to human referents,[5] and thus were coded as human for the sake of our analysis.

4 Wordless story books included ones by O'Shannessy (2004), as well as *Leaf* (King 2010), *Chicken and Cat* (Varon 2006) and *Hank Finds an Egg* (Dudley 2013). Videos included episodes of *Pingu*, *Tom & Jerry* and clips from Charlie Chaplin and Buster Keaton films.

5 For example, they were more likely to be elided, and were relatively more likely to appear clause-initially even when they were patients. These results are discussed below.

Results

A summary of word orders found in each type of study is shown in Table 16.2.

Table 16.2: Summary of word orders produced in experimental and spontaneous data.

	Experimental data		Spontaneous data	
SOV	957	51.31%	181	16.31%
SVO	415	22.25%	99	8.92%
SV	158	8.47%	184	16.58%
OSV	92	4.93%	83	7.48%
OVS	75	4.02%	23	2.07%
OV	113	6.06%	233	20.99%
VSO	2	0.11%	3	0.27%
VOS	1	0.05%	–	0%
VS	8	0.43%	24	2.16%
VO	9	0.48%	55	4.95%
V	35	1.88%	225	20.27%
Total	1865	100%	1110	100%

In the experimental data, we found all logically possible orders of V, S and O. Examples of some of these word orders are shown in (2). Each speaker produced on average 5.14 different orders (*SD* = 1.59, range = 1–8). The most frequent order was SOV (51.31 per cent), and SVO was the next most frequent (22.25 per cent). The least frequent were verb initial orders VSO and VOS, with only 2 and 1 observations, respectively. The results show a high degree of variability, given that these pictures are shown in an experimental context where there is no variation in the discourse context.

(2)a. *wati* *kutju-ngku* *antipina* *ila-n̪i*
man one-ERG fish(ACC) pull-PRS
S O V
'A man is pulling a fish.' (20180926-PJ03)

b. *minyma* *kutju-ngku* *ila-n̪i* *ngin̪taka*
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.' (20180930-PJ10)

h. *witi-ni*
catch-PRS
V
'(A woman) is catching (lizards).' (20181007-PJ35)

In the spontaneous data, all logically possible orders were produced as well, with the exception of VOS. Speakers on average produced 7.48 different orders (*SD* = 1.68, range = 5–10), although it should be noted these figures are not directly comparable to the experimental data, as some speakers were only represented in the data by a very small number of clauses. While the most frequent order, SOV, occurred 51.31 per cent of the time in the experimental data, only 16.31 per cent of clauses in the spontaneous data

had SOV order. The most frequent order OV was not much higher, at just 20.99 per cent. Note that we treat sentences produced with ellipsis as separate orders; while one could speculate that OV is underlyingly another order, such as SOV, we have no way of knowing which order is the underlying one in any particular instance; it could also be OVS or even OSV, for example. While the rates of ellipsis were much higher in the spontaneous data, the relative frequency of SVO and SOV was similar in both genres, with SOV approximately twice as frequent as SVO. This is discussed below. Overall, the spontaneous data were much more variable than the experimental data, largely due to this higher rate of ellipsis (57.58 per cent vs 19.2 per cent in the experimental data), as we detail further below. One further notable difference between the two datasets is the relative frequency of OSV. Putting aside the issue of ellipsis, and looking only at sentences produced with both S and O, OSV constitutes 5.97 per cent in the experimental data. However, in the spontaneous data, OSV constitutes 21.34 per cent of orders produced where both S and O arguments were present.

Choice of different word orders

Experimental data

We now look at what factors condition the choice of one word order over another, including the choice to omit one or both arguments. We first look at the experimental data, which allows us to see several factors at play, separate from the information packaging requirements of ongoing discourse. In particular, we can look at the role of humanness and grammatical function, as human and non-human characters were fully crossed in terms of their semantic role (agent or patient, expressed as subject and object in our data). Table 16.3 shows the word orders produced in the experimental data according to the humanness of the subject and object.[6]

6 Note that numbers of clauses that were eventually included are not exactly even, even though the stimuli were balanced for humanness – it seems that speakers may have been less likely to use typical transitive clauses to describe events with less prototypical character combinations; i.e. non-human S, human O sentences had the most exclusions due to a non-transitive clause being produced (only 397 remaining out of 588 possible, 67.52%), while human S, non-human O sentences had the lowest number of exclusions (537/588, 91.33%).

Table 16.3: Word orders produced in experimental data according to combinations of humanness and grammatical function.

SOV	SVO	SV	OSV	OVS	OV	VSO	VOS	VS	VO	V	Total
Human S, human O											
199	83	62	15	17	49	1	1	5	6	31	469
42%	18%	13%	3%	4%	10%	0%	0%	1%	1%	7%	100%
Human S, non-human O											
332	89	9	14	34	54	1	0	0	2	2	537
62%	17%	2%	3%	6%	10%	0%	0%	0%	0%	0%	100%
Non-human S, human O											
166	107	60	45	14	1	0	0	3	0	1	397
42%	27%	15%	11%	4%	0%	0%	0%	1%	0%	0%	100%
Non-human S, non-human O											
260	136	27	18	10	9	0	0	0	1	1	462
56%	29%	6%	4%	2%	2%	0%	0%	0%	0%	0%	100%
Total											
957	415	158	92	75	113	2	1	8	9	35	1865
51%	22%	8%	5%	4%	6%	0%	0%	0%	0%	2%	100%

Note: Some cells have been rounded down to 0%.

Hypothetically, if all word order variability in Pitjantjatjara were due to discourse-pragmatic factors, we might expect entirely SOV word orders in the experiment (if it is the syntactically basic word order), or all orders to be equally frequent (if word order is completely free and randomly determined). Neither of these possibilities is borne out by the data. Rather, patterns of word order choice in the experimental data seem to be influenced by the following four principles.

(1) Subject before object

There is a strong preference for subjects to appear before objects. SOV (3) and SVO (4) orders constitute the majority of responses in all character combinations, while OSV and OVS are a small minority. Overall, there were 1,374 (73.57 per cent) subject-before-object sentences, compared to 168 (9.01 per cent) object-before-subject sentences; the remaining 323 sentences had one or both arguments elided (17.32 per cent).

(3) *wati-ngku* *kaḻi* *waṉi-ngu*
man-ERG boomerang throw-PST
'The man threw a boomerang.' (20181002-PJ15)

(4) *wati* *kutju-ngku* *waṉi-nyi* *kaḻi*
man one-ERG throw-PRS boomerang
'The man threw a boomerang.' (20180926-PJ03)

(2) Human before non-human

There is a preference for human participants to appear before non-human ones, and this can sometimes conflict with (1) (although this is not often the case in naturalistic speech, as we outline below). That is, the subject-before-object principle may sometimes be violated where the object is human and the subject is non-human. Of the 92 OSV responses, 45 (48.9 per cent) were in the non-human:human condition, as in (5). However, OVS responses follow a different pattern; 34 of 75 OVS responses (45 per cent) had a human subject and a non-human object, while only 14 (18.67 per cent) had a non-human subject and a human object. In several of these responses, closer inspection shows that the subject was produced following a prosodic boundary and a pause, suggesting an 'afterthought' construction, where the speaker may have initially intended to produce an OV sentence, then decided to add the agent later, as illustrated in (6). This is in line with Bowe's analysis described above (1990, 113).

(5) *nyiṯayira* *kutjara* *tjuḻpu-ngku* *nguḻu-tjinga-ṉi*
boy two bird-ERG afraid-CAUS-PRS
'The bird scared two boys.' (20180927-PJ05)

(6) *tjiipi* *pampu-ṉu,* *pampa-ngku*
sheep touch-PST old.woman-ERG
'The old woman patted a sheep.' (20180930-PJ11)

(3) Avoid initial verbs

Regardless of the order of participants, there is a very strong preference that verbs do not appear sentence initially. There are only 49 sentences coded as verb initial in the experimental data (2.63 per cent), of which 33 are V-only. The verb appears before a core participant in just 16 responses (0.86 per cent). This preference is, in fact, even stronger than it would appear by these figures only. Our coding scheme is concerned with the order of subject,

object and verb, but, in fact, many of these 'verb initial' sentences have an adjunct appearing in sentence-initial position, such as the instrument in example (7). This is the case for 27 of those 49 'verb initial' sentences; therefore, only 22 responses (1.18 per cent) have a verb as the actual first element.

(7) *nyiila-ngka* *waka-n̲i* *tjitja-ngku*
needle-LOC pierce-PRS nurse-ERG
'The nurse is injecting (the child) with a needle.' (20181002-PJ17)

(4) Humans can be elided

Arguments can be freely elided in Pitjantjatjara, even in the experimental setting where there is no continuity in reference between clauses (i.e. the elided argument cannot be recovered from any preceding discourse, although may be visually present). Ellipsis of human arguments is much more likely than non-human arguments. The majority of SV responses had human patients (122 of 158, or 77.22 per cent), as in (8), and the majority of OV responses had human agents (103 of 113, 91.15 per cent), as (9). Of the 35 verb-only responses with no explicit mention of the agent or patient, 31 (88.57 per cent) of these were humans acting on humans; e.g. (10).

(8) *putji-ngku* *piri-ni*
cat-ERG scratch-PRS
'The cat is scratching (a girl).' (20180930-PJ11)

(9) *tjuki* *riitja-mila-n̲u*
chicken race-LOAN-PST
'(A woman) chased a chicken.' (20181007-PJ40)

(10) *kantu-n̲u*
kick-PST
'(A solder) kicked (a boy).' (20181007-PJ35)

Table 16.4 shows rates of NP ellipsis and NP expression in the experimental data according to humanness and grammatical function. Chi-square analyses showed that there is a significant relationship between humanness and ellipsis (χ^2 = 197.24, df = 1, p <.001), but not between grammatical function and ellipsis (χ^2 = 2.03, df = 1, p =.154). See below for further discussion of ellipsis, humanness and grammatical function in the spontaneous data.

Table 16.4: Ellipsis of arguments in experimental data according to humanness and grammatical function.

		Elided		Mentioned		Total
Human	Subject	144	14.31%	862	85.69%	1,006
	Object	162	18.71%	704	81.29%	866
Non-human	Subject	13	1.51%	846	98.49%	859
	Object	39	3.9%	960	96.1%	999
Total		368	9.6%	3,372	90.4%	3,730

From these four principles, it follows that SOV and SVO are the most frequent orders (a combined 1,372 of 1,865, 73.57 per cent), and that VOS is the least preferred order, with just a single observation. However, these four principles do not account for all variation in the data. There is still a substantial amount of individual preference, or even randomness, mediating these four principles. Some variation, such as the choice between SOV and SVO, cannot be accounted for by these principles at all. Detailed statistical analysis of word orders produced in the experimental data according to role and conceptual accessibility (i.e. humanness) can be found in Kidd et al. (2025).

Spontaneous data

We now compare our findings in the spontaneous data. As already mentioned, the most obvious difference between the two types of study is that there is more variability in the spontaneous data, largely due to much more frequent ellipsis.

Unlike the experimental data, the spontaneous data were not balanced for character humanness and, thus, it reflects more prototypical combinations of humanness and grammatical function than the experimental data. As seen in Table 16.5, human subjects acting on non-human objects account for 65.49 per cent of transitive clauses, while non-human subjects acting on human objects account for just 4.25 per cent. Nine sentences were excluded from this table where the identity, and therefore humanness, of an argument was unclear.

Table 16.5: Frequency of humanness/grammatical roles in spontaneous data.

	Human object	Non-human object	Total
Human subject	270 (24.52%)	721 (65.49%)	991 (90.01%)
Non-human subject	63 (5.72%)	47 (4.27%)	110 (9.99%)
Total	333 (30.25%)	768 (69.75%)	1101 (100%)

Looking at word order choices when both arguments are expressed with a noun phrase (i.e. not with a clitic, and not including ellipsis), we see in Figure 16.3 that most character combinations have quite similar distributions of word orders, except for non-human subjects acting on human objects, the least prototypical combination, shown to the right of Figure 16.3. With this combination, SVO is less common, while orders that front the human object (OSV and OVS) are used more. SOV, however, is still used often as well. The higher occurrence of object-initial orders for non-human:human clauses was also seen in the experimental data. While topicality may be a factor (if humans are more likely to be topics and topics are more likely to come first), humanness itself likely also plays a role. One caveat, however, is that this non-prototypical clause type only occurred 23 times in the spontaneous data; this is one reason why the complementary use of the experimental data is beneficial.

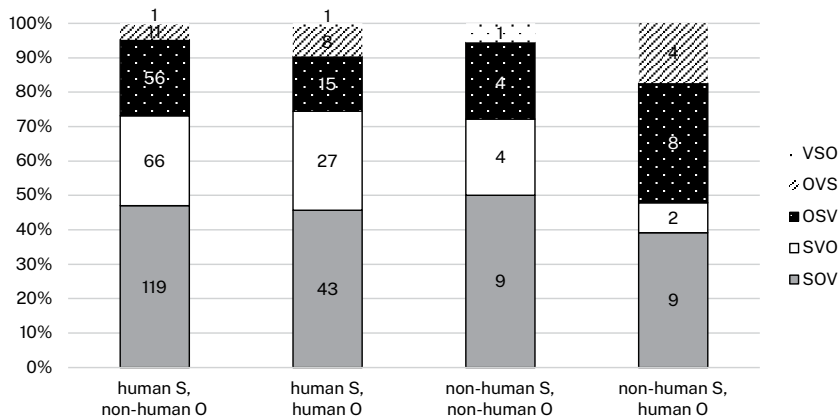

Figure 16.3: Word order frequency for combinations of human and non-human subjects and objects in spontaneous data, where both arguments are overtly expressed.

As in the experimental data, there is a certain amount of variability that cannot be explained by the four principles outlined above, or by information structure. We can see this clearly in some of the spontaneous data. For example, one of the storybooks includes a scene where the main character makes a ladder. Theoretically, each telling of this story by different speakers would have the same considerations of what is given or new, what is the topic or focus. However, there is substantial variation in how this particular event is expressed through different word orders; four different orders are attested in the 10 narrations of this story, as seen in (11)–(15) with some preceding and following utterances for context. In (13), where the speaker uses an SVO order, there is a pause between the verb and the object. This prosodic pattern in SVO sentences is associated with 'afterthoughts' in Bowe's (1990) and Goddard's (1985) analyses, although 'afterthought' may not be an appropriate description of this example, where the speaker has been talking about the ladder all along. The verb *palya-ṉu* 'do/make/fix-PST' is very vague, and it is the ladder itself that is the more prominent information, which then goes on to feature in several subsequent events. In (15), we see the speaker first using an SOV order in a quite elaborated sentence. She then restates the main substance of this episode – the making of the ladder – with an OSV order, where the object (the ladder) is topicalised as the most 'prominent' information in this event.

(11) CM: SOV

munu paluṟu puṉu puḻkanya uraṉu, tjaṟu tjunu munu paluṟu puṯu katuṉingi

'And he got a big piece of wood, put (it) down, and he couldn't lift (the egg) up.'

munu	***palu-ṟu***	***lata***	***kuḻupa***	***palya-ṉu***
and.SS	3SG.ERG	ladder	small	make-PST

'And he made a little ladder.'

ngampu palunya tjunkunytjikitjangku, manngu katu maḻakungku

'In order to put the egg back up in the nest.' (SW20190331-06-EggStory)

(12) TL: OV (following some overlapping and inaudible speech)

lata kuḻupa palya-ṉu

ladder small make-PST

'(He) made a little ladder.'

munu katira, lipula tjunu munu puṯu tatiṟa puṯu tjunkula

'And (he) took (it) and put (it) straight and couldn't climb (it), couldn't put (it).' (SW20190425-02-EggStory)

(13) CB: SVO (with object following pause)

uṉṯuṟa uṉṯuṟa katingu puṉu panyatjakutu, manngukutu panya

'He was pushing and pushing (it) to that tree, towards the nest.'

munu ***palu-ṟu*** *walytja-ngku* ***palya-ṉu,*** ***lata*** ***kulupa***

and.SS 3SG.ERG by.oneself-ERG make-PST ladder small

'And he made a little ladder on his own.'

puṉu tjuṯa kaṯaṟa lata tjunu

'He cut up sticks and put together a ladder.' (20190427-03-EggStory)

(14) AW: OSV

puṉu kutjara maṉkurpalta paluṟu uraṉu, lata-lata kuḻupa paluṟu one two three

'And then he got two, three pieces of wood. A little toy ladder, one, two, three.'

lata kuḻupa palu-ṟu palya-ṉu

ladder small 3SG.ERG make-PST

'He made a little ladder.'

lata palatja wati-liinamilaṉu, ngampu panya tjaṉangka nyaangka tjunu

'(He) leaned that ladder across, put that egg in the thing on his back.' (SW20190602-03-EggStory)

(15) AR: SOV, repeated as OSV

munu nyara paluṟu Hank, Hank-tu puṉu tjuṯa mantjinu

And that one Hank, Hank got lots of sticks.

munu	*palu-ṟu*	*lipula-ṟa*	*lipula-ṟa*	***palu-ṟu***	***lata***	***kuḻupa***	***palya-ṉu***
and.SS	3SG.ERG	level-MV	level-MV	3SG.ERG	ladder	small	make

tati-ntji-kitja-ngku	*ngampu*	*palu-nya*	*tju-nku-kitja-ngku*
climb-NMLZ-INTENT-ERG	egg	3SG-ACC	put-AUG-INTENT-ERG

'And he was laying (it) out (aligning the branches correctly on the ground), he made a little ladder to climb up, to put the egg up in the tree.'

puṉu	*katu,*	***lata***	***kuḻupa***	***palu-ṟu***	***palya-ṉu***
tree	high	ladder	small	3SG.ERG	make-PST

'He made a little ladder.' (SW20190422-01-EggStory)

As already mentioned, ellipsis was much more frequent in the spontaneous data compared to the experimental data. The principle of humans being elided more often is also true, to an even greater extent as it is confounded with givenness: humans are more likely to be the topic of discourse, and therefore represent given information in subsequent clauses. This is also compounded by a switch reference distinction in conjunctions: *munu* 'and.ss' is used to introduce a main clause whose subject is the same as that of the previous main clause. This is illustrated in (16), where different object referents are each mentioned, but the human subject is only mentioned in the first of the sequence of clauses. While the ellipsis of the subject can be explained on the basis of givenness, the motivation for the choice of OV vs VO order is not clear.

(16)

one,	***two,***	***three***	***step***	***palu-ṟu***	***nyaa-ṉu***	*four*
one	two	three	step	3SG.ERG	do.what-PST	four

'He took one, two, three steps, four steps.' (OSV)

munu	***tuwa***	***kutjup***	***aḻa-ṉu,***
and.SS	door	another	open-PST

'And opened the other door.' (OV)

munu ***ma-tjarpa-ngu***
and.ss away-enter-PST
'And went inside.' (V)

munu *pi̱ruku,* *watarku* ***wati-nya-nga-ngi*** ***kutjupa*** ***saita*** ***munu***
and.ss again unawares across-see-AUG-PST.IPFV other side and.ss
'And again he was looking across at the other side, not realising.' (VO)

munu ***nya-ngu*** ***Tiger-nya,*** ***ngalya-nya-nga-nyi***
and.ss see-PST Tiger-ACC hither-look-AUG-PRS
'And saw the tiger, looking this way.' (VO, V) (SW20190412-06-ChaplinLion)

The data (summarised in Table 16.6) were analysed using a log-linear analysis, which allows for the analysis of a 3-way contingency table in a similar manner to chi-squared analyses for 2 × 2 contingency tables.[7] The analysis provides several outputs, including the 3-way interaction term between all variables, 2-way interactions collapsing across the influence of the third variable (e.g. the interaction between Humanness and Ellipsis collapsing across Grammatical Function), and 2-way interactions when the influence of the third variable is removed (e.g. the interaction between Humanness and Ellipsis with the influence Grammatical Function removed).

The results are presented in Table 16.7. All interactions in the model were significant. These results show that the variables of humanness, ellipsis and grammatical function are all highly interconnected in discourse in Pitjantjatjara, further emphasising the utility of experimental data, which controls for humanness and grammatical function. The fact that grammatical function is significantly correlated with ellipsis in the spontaneous data but not the experimental data suggests that it is not grammatical function itself that influences ellipsis (as suggested by Bowe 1990, 66–67), but that this may be an indirect measure of topicality. It is furthermore possible that the relationship between humanness and ellipsis is due to topicality; even in the experimental data, speakers may assume that given some event (such as

7 Richard Lowry, Log-Linear Analysis for an AxBxC Contingency Table', Vassarstats, vassarstats.net/abc.html.

a man cooking a pig), the human is the topic, especially if an event typically involves a human. Further research would be needed to entirely disentangle humanness from topicality, and to understand, for example, why in the human:human condition, we find 52 SV responses, 49 OV responses and 31 V responses.

Table 16.6: Ellipsis of arguments in spontaneous data according to humanness and grammatical function.

		Elided		Mentioned		Total
Human	Subject	471	47.24%	526	52.76%	997
	Object	123	36.94%	210	63.06%	333
Non-human	Subject	32	29.09%	78	70.91%	110
	Object	300	39.06%	468	60.94%	768
Total		926	41.94%	1282	58.06%	2208

Table 16.7: Results of log-linear analysis of 3-way contingency table showing interactions between humanness, ellipsis and grammatical function in spontaneous data.

	G^2	df	*p*
Humanness*Ellipsis*Grammatical Function	906.18	4	<.0001
Humanness*Ellipsis	11.48	1	0.0007
Humanness*Grammatical Function	877.06	1	<.0001
Ellipsis*Grammatical Function	10.08	1	0.0015
Humanness*Ellipsis (-Grammatical Function)	19.04	2	<.0001
Humanness*Grammatical Function (-Ellipsis)	884.62	2	<.0001
Ellipsis*Grammatical Function (-Humanness)	17.64	2	0.0001

Generational change?

Langlois (2004) describes a shift from a basic SOV order to SVO, which was attributed to increased contact with English among teenage speakers in the 1990s in Areyonga. There are several features which distinguish English word order patterns from Pitjantjatjara, features which may therefore be liable to change in intense contact or incipient language shift situations. One is that word order is used to determine grammatical relations in English, whereas case marking is used to determine grammatical relations

in Pitjantjatjara. Language contact in other parts of Australia has led to a shift from traditional argument marking systems (such as ergativity or head-marking on verbs) to ones where grammatical relations are primarily distinguished by phrase structure (Bavin and Shopen 1985; Lee 1987; McGregor 2002; Meakins and O'Shannessy 2010; Pensalfini 1999; Richards 2001; Schmidt 1985). There is no sign that ergative case marking is becoming optional in Pitjantjatjara, leading to a greater reliance on word order to distinguish grammatical relations, so we do not discuss this further.

Two further dimensions, often conflated, are word order variability vs rigidity and the use of SVO as the most common word order (where both subject and object are expressed). Shifting to a more rigid word order with a greater relative frequency of SVO may occur in language contact situations; however, these are potentially independent changes. A language may change to using SVO as its most frequent word order, while still maintaining an equal degree of flexibility; or the opposite may occur where a preferred non-SVO word order increases its frequency relative to other orders, resulting in a more rigid preference for a particular order. For this reason, we discuss these phenomena separately, first looking at variability vs rigidity itself, then looking at the relative frequency of SOV and SVO to see if there a potential change in the most common word order.

Word order variability

The application of information-theoretic techniques to large corpora allows linguists to precisely quantify variability across many different outcome variables, including word order flexibility (e.g. Levshina 2019). We can see more fine-grained differences in the degree of variation in a language, instead of categorical designations such as 'free', 'fixed', or 'flexible' (see Levshina et al. 2023). This is done by calculating the Shannon entropy H (Shannon 1948) of some probability distribution; we do so using the entropy package (Strimmer 2021) in R (version 4.0.3).[8] A dataset where 99 per cent of clauses are SOV and 1 per cent are SVO would have a very low entropy ($H = 0.081$ to be precise). If we were to record one more sentence, we would be quite certain that it will also be SOV, whereas we would be very surprised if it were SVO. The more variable the probability distribution, the higher the entropy; if there was a free word order language where the 11 possible word

8 *R: A Language and Environment for Statistical Computing*, v.4.0.3, R Core Team, R Foundation for Statistical Computing, 2020, www.R-project.org.

orders were equiprobable in some corpus, we would have a higher degree of uncertainty about what the next recorded sentence would be (H = 3.46). Thus, we can ask of our data: how variable is word order in Pitjantjatjara, where there is supposedly a basic word order but this doesn't determine grammatical relations? Are young people shifting to a more English-like distribution that is more rigid and predictable, as might be expected based on Langlois's observations and studies of other Australian languages in contact with English?

The first question is relatively straightforward to answer, with the only caveat being the difference between the two datasets. Previously, we noted that the spontaneous data appeared to be more variable than the experimental data. We can confirm this impression by calculating the entropy of both datasets. However, one issue with Shannon entropy is that comparison of entropy across distributions is not necessarily reliable where there are different numbers of observations, as it is not clear how well a sample of any given size is representative of the real world (see e.g. Back et al. 2019; Futrell et al. 2015; Levshina et al. 2023). In order to reliably compare the distributions across the two datasets (where n = 1,865 and n = 1,110, in the experimental and spontaneous data, respectively), we use the Chao-Shen method of calculating H throughout this chapter (Chao and Shen 2003). We do this using the entropy.ChaoShen() function within the entropy package (Strimmer 2021). This gives H = 2.11 for the experimental data, and H = 2.87 for the spontaneous data.[9] This confirms the descriptive observation that the spontaneous data are more variable than the experimental data. This is to be expected given that we expect much more ellipsis in spontaneous data where the preceding discourse context enables recovery of referents. Even putting aside this issue of ellipsis, we can also compare clauses with two fully expressed NPs only. For clauses with both S and O expressed as an NP, there is still a more unpredictable distribution in the spontaneous data (H = 1.81) compared to the experimental data (H = 1.42). This suggests that aside from the issue of ellipsis, the experimental setting leads Pitjantjatjara speakers,

9 Use of the Chao-Shen method in fact gives a very similar result (identical to two decimal points) as the standard method of calculating Shannon entropy. Levshina et al. (2023) find that when comparing word order variability with two possible outcomes (such as order of S/V or O/V) 'sample sizes of more than 300 clauses produce entropy gaps less than 0.1', and that this stabilised further after a threshold of 500 clauses. See also Back et al. (2019) for a method of determining the ideal sample size for calculating entropy with a given confidence level for a given number of possible outcomes (i.e. 11 logically possible word orders in our case).

on average, to choose less variable word orders than in spontaneous speech. One specific difference mentioned above is that OSV was more frequent in the spontaneous data than in the experimental data.

The second question of whether younger speakers have a less flexible word order is more complex. While we can calculate the variability of each individual speaker's productions and see whether there is a correlation with age, this is complicated by differences in the amount of data available per speaker. There are only 18 speakers in the corpus of spontaneous data, and some are overrepresented compared to others. Here, we instead use the experimental data, as there are 49 speakers, and each one has the same number of responses to the same stimuli (apart from exclusions detailed above). Certain considerations need to be taken in the interpretation of this data, however, as we detail below.

There is a correlation between age and word order entropy, but in the opposite direction to what would be expected if there was a shift to a more English-like system of word order rigidity. Younger speakers' productions are significantly more variable, that is, have a higher entropy, than older speakers' (M = 1.14, SD = .412, $r(47)$ = -.41, t = -3.05, d = -.89, p < .005). However, when gender is taken into account, we find that there is no significant correlation between age and entropy among the male participants, while it is significant for the female participants. This is shown in Figure 16.4. Each individual is plotted with a dot – black for women and grey for men – with entropy on the y-axis and age on the x-axis. The lighter grey trendline shows that variability among men shows a small but non-significant increase with age ($r(17)$ = .3, t = 1.41, p = .179), such that there is no evidence to suggest there is meaningful change across generations. In contrast, the black trendline shows that younger women have more variable productions than older women, as revealed by a strong negative correlation between entropy and age ($r(28)$ = -.67, t = -4.78, p < .001). The younger women seem to be behaving similarly to men of their age, however, suggesting that the effect may be driven by more word order rigidity in the productions of the older women.

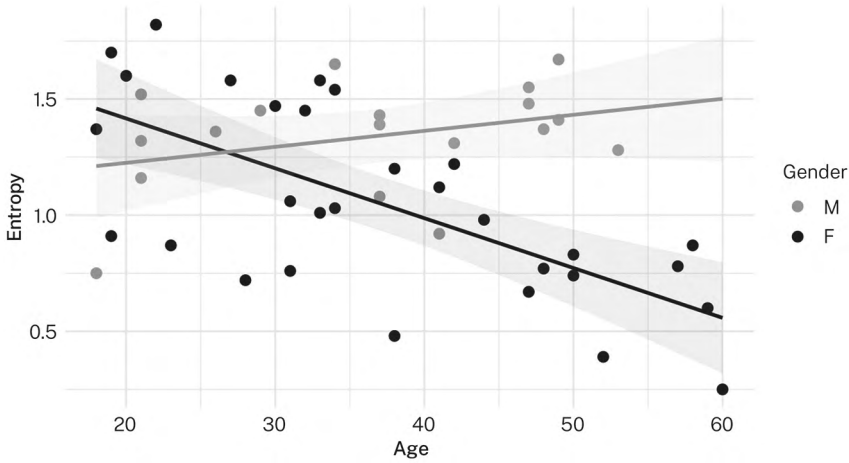

FIGURE 16.4: Individual levels of variability in word order choice in experimental data, measured as entropy, by age and gender.

Many of the older women who participated in the experiment have, for many decades, worked as teachers, interpreters, or translators. They are considered language experts in the community and feel an obligation to keep language strong. They may therefore hold more prescriptive views about what is 'correct' Pitjantjatjara; see Wilmoth (2022) for more detailed discussion of these language ideologies. These participants also happened to be those who produced overwhelmingly SOV responses in the experimental study. There were four speakers who contributed to both the experimental data and the spontaneous data. Their productions are shown in Figure 16.5, with the proportion of orders for each individual in the experimental data on the left of the figure and in the spontaneous data on the right. Although this is a small number of individuals, we can clearly see that the older women (AW, MD, UT, born 1968, 1959 and 1966 respectively) produced drastically different distributions to their more everyday speech; however, the younger woman (SM, born 1999), had relatively similar levels of variability in both genres.

This supports our view that this apparent trend may be driven by these older women's approach to the experimental task, rather than an increase in variability among the younger generations. No straightforward explanation from language contact or regular 'system-internal' change would account for an increase in variability among younger people; rather, this apparent trend may be a result of these older women's prescriptive ideals manifesting in less variability in an experimental setting. We therefore conclude that word order variability itself appears to be stable across generations.

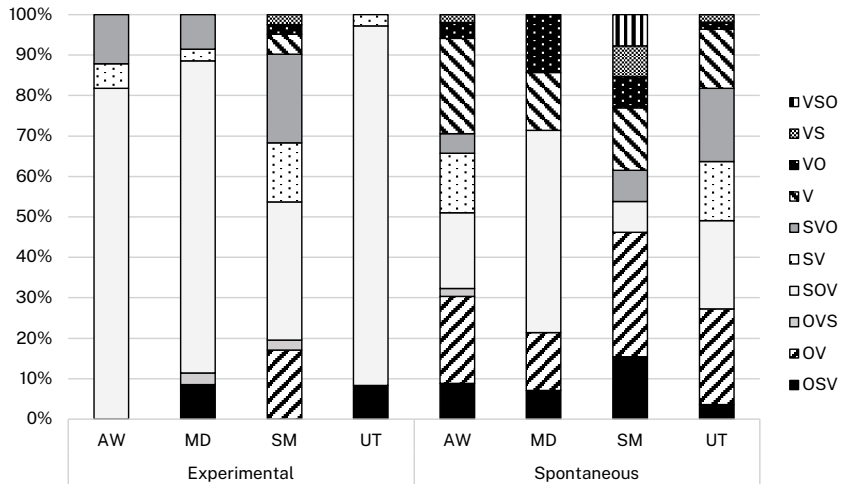

Figure 16.5: Word order choices by four individuals (AW, MD, SM, UT) in the experimental setting vs spontaneous data.

Most common word order

We now turn to the question of whether the most common word order in Pitjantjatjara (where both subject and object are expressed) is shifting from SOV to SVO as a result of contact with English, as argued by Langlois (2004). It should be noted that our data were recorded in a different time and place from Langlois's, and so differences in our results do not necessarily invalidate her findings. Here again, we compare the experimental findings compared to the spontaneous data. An analysis of word order choice (SVO vs SOV) and age (continuous measure, range 18–60 years old) was conducted with logistic mixed effects models in R (R Core Team 2019). We found a significant effect of age and word order choice in the experimental data. Younger participants were more likely to produce SVO sentences compared to the older population (β = -0.07, *SE* = 0.02, 95% CI [-0.10, -0.03]). In contrast, there was no significant effect of age and the choice to produce SVO sentences (compared to SOV) in the spontaneous data (β = 0.01, *SE* = 0.01, 95% CI [-0.02, 0.01]).[10] The experimental data are shown on the right-hand side of Figure 16.6, where we see a clear increase in the relative frequency of SVO compared to SOV with each successively younger age group. Speakers aged 50 and above produced 236 SOV responses (81.94 per cent), compared to 52 SVO responses (18.06 per cent), while this split

10 Linear regression model that evaluated word order choice (SVO vs SOV) and age (range 17–69).

was almost exactly even among those 18–29 years old (with 203 SOV and 198 SVO responses, or 50.62 per cent and 49.38 per cent respectively). However, there is no such trend in the spontaneous data, as seen in the right-hand side of Figure 16.6 (note that the middle two groups have a lower number of tokens due to the study design). The relative frequency of SOV vs SVO is stable across age groups, with SVO comprising 32.73 per cent of SVO/SOV orders among the oldest group (74 SVO vs 36 SOV) and 36.76 per cent among the youngest group (50 SVO and 86 SOV). As discussed above, one complicating factor in the experimental data is that speakers, particularly older women, may have been producing what they view as correct Pitjantjatjara, namely SOV. Assuming the corpus data more closely reflects everyday usage, it is not clear that there is any major change to the language's most common word order.

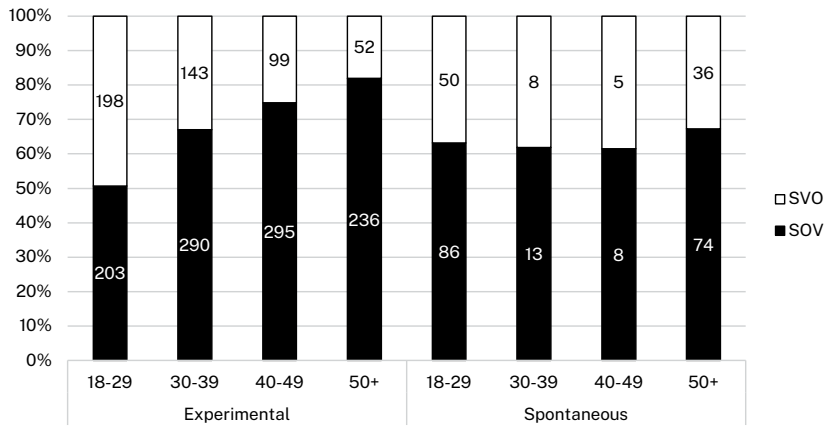

Figure 16.6: Relative proportion of SOV and SVO orders in experimental and spontaneous data according to age.

Conclusion

This chapter investigated word order variability in Pitjantjatjara. In contrast to other literature on the topic in Pitjantjatjara and other Australian languages more broadly, we use a combination of experimental and spontaneous data. This has allowed us to isolate humanness and grammatical function from pragmatic factors such as information structure and prominence. Previous descriptions of word order in Pitjantjatjara and Yankunytjatjara have argued for a basic SOV order, with specific word orders arising due to topicalisation,

focus, prominence, afterthoughts and so on (Bowe 1990; Goddard 1985; Rose 2001). Langlois (2004) also argues that there has been a shift among teenage speakers to a basic SVO order due to contact with English. We find little evidence to suggest that there is a single underlying or 'basic' word order, with variability across all speakers and both datasets that cannot always be attributed to discourse factors such as topicalisation or afterthought, although subject initial orders are certainly more common. We also find very little evidence of change in these patterns in younger speakers.

We discussed some of the factors driving different word order choices. We find four general principles that hold in both datasets, and these can be deduced even more clearly due to the inclusion of the experimental data, as this is more balanced and not confounded by variation in the discourse status of referents.

1. Subject before object
2. Human before non-human
3. Avoid verb initial
4. Humans can be elided.

These principles guide some, but not all, of the variation in word order. While the spontaneous data also has clear examples of word order differences attributable to topicalisation, focus, or afterthoughts, as in previous descriptions, there is also a degree of randomness and individual preference. Having established these principles, further research is needed to understand how these interact with discourse-structural features such as accessibility, topicality and focus.

We also investigated the hypothesis that there is a shift to a more rigid SVO basic word order due to contact with English. Unlike previous studies of this phenomenon in Australian languages (e.g. Schmidt 1985), we explicitly separated this into two dimensions of potential change: word order rigidity vs flexibility as measured by entropy, and SOV vs SVO as the most common word order. In terms of word order variability itself, the picture painted by the two datasets is complex, but we find that variability appears to be stable, if not even increasing among young people. Likewise, while the experimental data show a trend towards a more equal balance between SOV and SVO among younger speakers, compared to older speakers who use predominantly SOV over SVO, the spontaneous data show that the relative frequency of the two orders is actually stable. Assuming this more closely reflects people's everyday speech, it appears that at least in Pukatja, there is

no shift to SVO. Both of these findings are contrary to expectations and suggest the fundamentals of Pitjantjatjara grammar are not shifting due to language contact (see Wilmoth 2022).

Throughout, we have also shown the benefits of using experimental data. This was useful for balancing combinations of humanness and grammatical function, as certain combinations (such as a non-human agent acting on a human patient) occur very infrequently in more naturalistic speech. The lack of variation in discourse status means that we can see some level of 'baseline' preferences in the language, and separate syntactic and semantic factors from pragmatic ones. These factors, such as humanness, have rarely been discussed in previous research on word order variability, which has mostly focused on discourse factors. The use of experimental data also allows us to clarify that a 'basic' SOV word order is best understood to mean 'most frequent' in Pitjantjatjara and not 'syntactically underlying' or 'pragmatically unmarked', as pragmatic factors were controlled in the experimental study, yet most speakers produced a wide range of word orders. We also showed that ellipsis is significantly associated with humanness in the experimental data, but not grammatical function. This suggests that the association with grammatical function in the spontaneous data is due to subjects often being topics or otherwise given, and thus tend to be elided on a pragmatic basis, not due to the grammatical relation of subject *per se*. The use of experimental data also ensures comparability across speakers and situations, since all speakers describe all pictures. However, care must be taken in interpreting these results and comparing them to the spontaneous data. In particular, we found that certain older female participants appeared to have a different approach to the task, likely due to their experience with language teaching and translating, and their responsibility to uphold 'correct' Pitjantjatjara. Thus, as with any form of experimental data, what one gains in experimental control can come at the cost of a degree of ecological validity, highlighting the importance of considering data from multiple genres, as we have in this study.

In this chapter, we have shown that the factors governing word order in Pitjantjatjara are almost as varied as the word order options available. As Simpson (1983, 1991) argued for Warlpiri, it is clear that word order does not play a role in determining grammatical functions in Pitjantjatjara, and that the syntax does not prescribe a particular word order for transitive clauses. Furthermore, in spontaneous discourse, it appears that pragmatic considerations such as topicality and givenness or newness of information may determine word order choices, similarly to what Simpson (2007) has

argued for Warlpiri. However, through investigation of the experimental data, we find other factors also to be in play, including the humanness of subjects and objects and the prescriptive choices of individual speakers. These findings support the claim that word order is not determined by syntactic principles alone, but is able to be varied by speakers under different linguistic conditions, and to express factors such as the humanness of agents and patients, the prototypicality of subject-object combinations, as well as information and discourse structure. We hope that this research will highlight the interesting possibilities of exploring word order choices in these flexible word order languages and encourage other researchers to undertake detailed studies such as this across multiple genres and corpora.

Acknowledgements

Jane Simpson's work on morphosyntax, word order and discourse structure in Warlpiri has been an important predecessor of this work, and we can only hope to emulate the careful, insightful analysis and thoughtful representation of the language in its own terms, which is the trademark of her research. She has also been an encouraging and supportive colleague for all of us, in different ways, over many, many years, and we will be forever grateful for her positivity and intellectual generosity. We are grateful for the many Pitjantjatjara speakers who have contributed to this study, particularly Sandra Lewis, for her assistance and guidance. We are also grateful to the editors and anonymous reviewers for their feedback that helped to improve this chapter. This research was made possible with funding from the Australian Research Council (DP210102836, CE140100041), the Max Planck Gesellschaft, the University of Melbourne and an Australian Government Research Training Program Scholarship.

References

Austin, Peter. 2001. 'Word Order in a Free Word Order Language: The Case of Jiwarli'. In *Forty Years on: Ken Hale and Australian Languages*, edited by Jane Simpson, David Nash, Mary Laughren, Peter Austin and Barry Alpher, 305–24. Canberra: Pacific Linguistics. doi.org/10.15144/PL-512.305.

Austin, Peter and Joan Bresnan. 1996. 'Non-Configurationality in Australian Aboriginal Languages'. *Natural Language & Linguistic Theory* 14, no. 2: 215–68. doi.org/10.1007/BF00133684.

Back, Andrew D., Daniel Angus and Janet Wiles. 2019. 'Determining the Number of Samples Required to Estimate Entropy in Natural Sequences'. *IEEE Transactions on Information Theory* 65, no. 7: 4345–52.

Bavin, Edith L. and Tim Shopen. 1985. 'Warlpiri and English: Languages in Contact'. In *Australia, Meeting Place of Languages*, edited by Michael Clyne, 81–94. Canberra: Pacific Linguistics. doi.org/10.15144/PL-C92.81.

Bowe, Heather J. 1990. *Categories, Constituents, and Constituent Order in Pitjantjatjara: An Aboriginal Language of Australia*. London: Routledge.

Chao, Anne and Tsung-Jen Shen. 2003. 'Nonparametric Estimation of Shannon's Index of Diversity When There Are Unseen Species in Sample'. *Environmental and Ecological Statistics* 10, no. 4: 429–43. doi.org/10.1023/A:1026096204727.

Defina, Rebecca, Catalina Torres and Hywel Stoakes. 2020. 'Scaling Processes of Clause Chains in Pitjantjatjara'. In *Interspeech 2020*, 1868–72. ISCA. doi.org/10.21437/Interspeech.2020-2101.

DITRDC, AIATSIS and ANU. 2020. *National Indigenous Languages Report*. Canberra: Australian Government.

Dudley, R. 2013. *Hank Finds an Egg*. White Plains: Peter Pauper Press.

Futrell, Richard, Kyle Mahowald and Edward Gibson. 2015. 'Large-Scale Evidence of Dependency Length Minimization in 37 Languages'. *Proceedings of the National Academy of Sciences* 112, no. 33: 10336–41. doi.org/10.1073/pnas.1502134112.

Goddard, Cliff. 1985. *A Grammar of Yankunytjatjara*. Alice Springs: IAD Press.

Hale, K. 1981. 'Preliminary Remarks on the Grammar of Part-Whole Relations in Warlpiri'. In *Studies in Pacific Languages and Cultures in Honor of Bruce Biggs*, edited by Jim Hollyman and Andrew Pawley, 333–44. Auckland: Linguistic Society of New Zealand.

Hale, Ken. 1983. 'Warlpiri and the Grammar of Non-Configurational Languages'. *Natural Language & Linguistic Theory* 1, no. 1: 5–47. doi.org/10.1007/BF00210374.

Kidd, E., Gabriela Garrido Rodríguez, S. Wilmoth, J. E. Garrido Guillén and R. Nordlinger. 2025. 'How Does Speaking A Free Word Order Language Influence Sentence Planning and Production? Evidence From Pitjantjatjara (Pama-Nyungan, Australia)'. *Cognitive Science* 49, no. 7: e70087. doi.org/10.1111/cogs.70087.

King, S. M. 2010. *Leaf*. Gosford: Scholastic Press.

Langlois, Annie. 2004. *Alive and Kicking: Areyonga Teenage Pitjantjatjara*. Canberra: Pacific Linguistics.

Lee, Jennifer R. 1987. *Tiwi Today: A Study of Language Change in a Contact Situation*. Canberra: Pacific Linguistics.

Levshina, Natalia. 2019. 'Token-Based Typology and Word Order Entropy: A Study Based on Universal Dependencies'. *Linguistic Typology* 23, no. 3: 533–72. doi.org/10.1515/lingty-2019-0025.

Levshina, Natalia, S. Namboodiripad, M. Allassonnière-Tang, M. Kramer, L. Talamo, A. Verkerk, S. Wilmoth, G. Garrido Rodriguez, T. M. Gupton, E. Kidd, Z. Liu, C. Naccarato, R. Nordlinger, A. Panova and N. Stoynova. 2023. 'Why We Need a Gradient Approach to Word Order'. *Linguistics* 61, no. 4: 825–83. doi.org/10.1515/ling-2021-0098.

McGregor, William. 2002. 'Structural Changes in Language Obsolescence: A Kimberley (Australia) Perspective'. *SKY Journal of Linguistics*, no. 15: 145–85.

Meakins, Felicity and Carmel O'Shannessy. 2010. 'Ordering Arguments about: Word Order and Discourse Motivations in the Development and Use of the Ergative Marker in Two Australian Mixed Languages'. *Lingua* 120, no. 7: 1693–713. doi.org/10.1016/j.lingua.2009.05.013.

Naessan, Petter. 2008. 'Some Tentative Remarks on the Sociolinguistic Vitality of Yankunytjatjara in Coober Pedy, South Australia'. *Australian Journal of Linguistics* 28, no. 2: 103–38. doi.org/10.1080/07268600802308741.

Norcliffe, Elisabeth, Agnieszka E. Konopka, Penelope Brow and Stephen C. Levinson. 2015. 'Word Order Affects the Time Course of Sentence Formulation in Tzeltal'. *Language, Cognition and Neuroscience* 30, no. 9: 1187–208. doi.org/10.1080/23273798.2015.1006238.

Nordlinger, Rachel. 1998. *Constructive Case: Evidence from Australian Languages*. Stanford: CSLI Publications.

Nordlinger, Rachel and Joan Bresnan. 2011. 'Lexical-Functional Grammar: Interactions Between Morphology and Syntax'. In *Non-Transformational Syntax: Formal and Explicit Models of Grammar*, edited by Robert D. Borsley and Kersti Börjars, 112–40. Hoboken: Wiley. doi.org/10.1002/9781444395037.ch3.

Nordlinger, Rachel, Gabriela Garrido Rodriguez and Evan Kidd. 2022. 'Sentence Planning and Production in Murrinhpatha, an Australian "Free Word Order" Language'. *Language* 98, no. 2: 187–220. doi.org/10.1353/lan.2022.0008.

O'Shannessy, C. 2004. *The Monster Stories: Picture Stimuli for Elicited Production*. Unpublished series. Max Planck Institute for Psycholinguistics, Nijmegen.

Pensalfini, Rob. 1999. 'The Rise of Case Suffixes as Discourse Markers in Jingulu – A Case Study of Innovation in an Obsolescent Language'. *Australian Journal of Linguistics* 19, no. 2: 225–40. doi.org/10.1080/07268609908599582.

R Core Team. 2019. 'R: A Language and Environment for Statistical Computing'. Manual. Vienna, Austria. www.R-project.org/.

Richards, Norvin. 2001. 'Leerdil Yuujmen Bana Yanangarr (Old and New Lardil)'. In *Forty Years On: Ken Hale and Australian Languages*, edited by Jane Simpson, David Nash, Mary Laughren, Peter Austin and Barry Alpher, 431–46. Canberra: Pacific Linguistics. doi.org/10.15144/PL-512.431.

Rose, David. 2001. *The Western Desert Code: An Australian Cryptogrammar*. Canberra: Pacific Linguistics.

San Roque, Lila, L. Gawne, D. Hoenigman, J. C. Miller, A. Rumsey, S. Spronck, A. Carroll and N. Evans. 2012. 'Getting the Story Straight: Language Fieldwork Using a Narrative Problem-Solving Task'. *Language Documentation & Conservation* 6: 135–74.

Schmidt, Annette. 1985. 'The Fate of Ergativity in Dying Dyirbal'. *Language* 61, no. 2: 378–96. doi.org/10.2307/414150.

Shannon, Claude E. 1948. 'A Mathematical Theory of Communication'. *The Bell System Technical Journal* 27, no. 3: 379–423. doi.org/10.1002/j.1538-7305.1948.tb01338.x.

Simpson, Jane. 1983. 'Aspects of Warlpiri Morphology and Syntax'. PhD diss., Massachusetts Institute of Technology.

Simpson, Jane. 1991. *Warlpiri Morpho-Syntax: A Lexicalist Approach.* Dordrecht: Springer.

Simpson, Jane. 2007. 'Expressing Pragmatic Constraints on Word Order in Warlpiri'. In *Architectures, Rules, and Preferences: Variations on Themes by Joan W. Bresnan*, edited by Annie Zaenen, Jane Simpson, Tracy Holloway King, Jane Grimshaw, Joan Maling and Chris Manning, 403–28. Stanford: Center for the Study of Language and Information.

Simpson, Jane and Ilana Mushin. 2008. 'Clause-Initial Position in Four Australian Languages'. In *Discourse and Grammar in Australian Languages*, edited by Ilana Mushin and Brett Baker, 25–58. Amsterdam: John Benjamins Publishing Company.

Strimmer, Jean Hausser and Korbinian Strimmer. 2021. 'Entropy: Estimation of Entropy, Mutual Information and Related Quantities'. CRAN.R-project.org/package=entropy.

Swartz, Stephen. 1991. *Constraints on Zero Anaphora and Word Order in Warlpiri Narrative Text*. SIL-AAIB Occasional Papers, no. 1. Darwin: Australian Aborigines and Islanders Branch, Summer Institute of Linguistics.

Tabain, Marija and Janet Fletcher. 2012. 'A Preliminary Study of Intonation in Pitjantjatjara'. In *Proceedings of the 14th Australasian International Conference on Speech Science and Technology*, edited by Felicity Cox, Katherine Demuth, Susan Lin, Kelly Miles, Sallyanne Palethorpe, Jason Shaw and Ivan Yuen, 213–16. Canberra: Australasian Speech Science and Technology Association (ASSTA).

Varon, S. 2006. *Chicken and Cat*. 1st ed. New York: Scholastic Press.

Wilmoth, Sasha. 2022. 'The Dynamics of Contemporary Pitjantjatjara: An Intergenerational Study'. PhD diss., University of Melbourne.