



Cross-clause planning in Nungon (Papua New Guinea): Eye-tracking evidence

Hannah S. Sarvasy¹ · Adam Milton Morgan² · Jenny Yu¹ · Victor S. Ferreira³ · Shota Momma⁴

Accepted: 10 October 2021
© The Author(s) 2021

Abstract

Hundreds of languages worldwide use a sentence structure known as the “clause chain,” in which 20 or more clauses can be stacked to form a sentence. The Papuan language Nungon is among a subset of clause chaining languages that require “switch-reference” suffixes on nonfinal verbs in chains. These suffixes announce whether the subject of each upcoming clause will differ from the subject of the previous clause. We examine two major issues in psycholinguistics: predictive processing in comprehension, and advance planning in production. Whereas previous work on other languages has demonstrated that sentence planning can be incremental, switch-reference marking would seem to prohibit strictly incremental planning, as it requires speakers to plan the next clause before they can finish producing the current one. This suggests an intriguing possibility: planning strategies may be fundamentally different in Nungon. We used a mobile eye-tracker and solar-powered laptops in a remote village in Papua, New Guinea, to track Nungon speakers’ gaze in two experiments: comprehension and production. Curiously, during comprehension, fixation data failed to find evidence that switch-reference marking is used for predictive processing. However, during production, we found evidence for advance planning of switch-reference markers, and, by extension, the subjects they presage. We propose that this degree of advance syntactic planning pushes the boundaries of what is known about sentence planning, drawing on data from a novel morpheme type in an understudied language.

Keywords Sentence processing · Language comprehension · Language production · Eye · movements · Syntactic processing

Introduction

Imagine a speaker telling a story, and upon describing the current action, they must announce in advance whether the next action will be done by the same actor (“NOSWITCH”) or instead will be done by a different actor (“SWITCH”). A simple story—I walked to the store. My friends were standing outside. They waved to me. I waved back. I did my

shopping, then came home—would sound something like: *I walked-SWITCH to the store. My friends were standing-NOSWITCH outside. They waved-SWITCH to me. I waved-NOSWITCH back. I did-NOSWITCH my shopping, then came home.* Is this difficult to do? Apparently not, judging by the ease with which it is done by native speakers of numerous indigenous languages of the Amazon, North America, and New Guinea.

This “switch-reference marking” (Haiman & Munro, 1983; van Gijn & Hammond, 2016) is intriguing from a language processing perspective. There is extensive literature on how speakers track relationships between words within a clause (e.g., agreement; Wagers et al., 2009), and where two elements in different clauses share a referent (e.g., long-distance dependencies; Clifton & Frazier, 1989). To our knowledge, however, there is no previous research on processing of a feature like switch-reference marking, where speakers must compute relations between distinct referents across different clauses.

✉ Hannah S. Sarvasy
h.sarvasy@westernsydney.edu.au

¹ The MARCS Institute for Brain, Behaviour and Development, Western Sydney University, Locked Bag 1797, Penrith, NSW 2571, Australia

² Department of Neurology, NYU School of Medicine, New York, NY, USA

³ Department of Psychology, University of California, San Diego, CA, USA

⁴ Department of Linguistics, University of Massachusetts, Amherst, MA, USA

To understand switch-reference marking, one must first understand the sentence type in which it occurs. In English (and other languages of Europe), clauses can be combined in one of two ways: coordination (i.e., use of conjunctions like *and*, and *or*, as in *The dog barked and the cat ran away*) or subordination (e.g., relative clauses, as in *The dog barked at the cat that ran away*). However, in a number of languages, including Japanese, Korean, Turkish, Tibetan, Chechen, and Burmese, there is a third way to combine clauses. In “clause chains” (Dooley, 2010; Longacre, 1985, 2007; Sarvasy, 2021), multiple clauses describing sequences of actions or events can be uttered one after another, forming a long sentence, as in (1), where brackets indicate clauses:

1. [*The cat biting the dog*], [*running under the table*], [*finding its bowl empty*], [*the dog still barking at it*], [*the cat fled outside*].
 ‘*The cat bit the dog. It then ran under the table, where it found its bowl empty. The dog was still barking at it. The cat then fled outside.*’

Clause chains may contain 20 or more clauses, yet only the very last verb conveys tense, while the rest appear in an un-tensed form. If the sentence lacks temporal adverbs such as *yesterday*, a listener must wait for the last verb to find out whether the sequence of events is construed as past, present, or future (Sarvasy, 2020), apparently presenting a processing challenge.

Among clause chaining languages, a subset (largely in Amazonia, North America, and New Guinea) requires speakers to announce in advance whether the subject of the following clause will be the same or different from the current subject, by way of a particular suffix (or other type of marker) on the verb. If English were a language with switch-reference marking, the example clause chain in (1) might look something like:

2. [*The cat biting-NOSWITCH the dog*], [*running-NOSWITCH under the table*], [*finding-SWITCH its bowl empty*], [*the dog still barking-SWITCH at it*], [*the cat fled outside*].

The present paper presents the first psycholinguistic investigations of switch-reference marking of which we are aware.

When listening to speech, it is generally agreed that sentences are processed incrementally (Altmann & Mirković, 2009) and predictively (DeLong et al., 2005). Comprehenders use various sources of information for prediction. For instance, Mitsugi (2017) showed that Japanese speakers use case morphology (markers of grammatical role: subject, object, indirect object, etc.) as cues for predictive processing. Similarly, Altmann and Kamide (1999) found that

comprehenders’ gaze travels more to a cake after hearing the verb *eat* but more to a ball after hearing the verb *move*. However, relatively little is known about how or whether morphological features on verbs (e.g., suffixes for agreement, tense, or switch-reference) are used to predictively guide comprehension (but see Pizarro-Guevara & Wagers, 2020). This is in part because in most well-studied languages, verbal morphology often does not contain clues to upcoming information. For instance, in English, verbs agree with subjects, so verbal morphology could in principle be used to predict the subject. But English verbs almost always come *after* the subject (although Lukyanenko & Fisher, 2016, show that in questions, where the English verb precedes the subject, number agreement inflection on verbs does aid prediction). To study prediction on the basis of verbal morphology, one needs morphological cues to *upcoming* information. Switch-reference markers therefore present a prime case for studying prediction on the basis of verbal morphology.

It is also important to validate the finding of predictive processing during comprehension in nonindustrialized populations who speak lesser-studied languages. Because most studies to date have relied on a certain type of participant (university students in industrialized nations), a finding of prediction based on switch-reference marking in a language of rural New Guinea would complement existing evidence from well-studied languages like English (DeLong et al., 2005), German (Kamide, Scheepers, et al., 2003b), and Japanese (Kamide, Altmann, et al., 2003a; Yoshida, 2004). Expanding the list of languages is important in establishing the generality of the claim that language processing operates predictively.

Switch-reference marking also has implications for research into language production, since speakers must know the subject of the next clause in order to produce switch-reference marking correctly. It is generally accepted that speakers plan speech in advance to some degree, although the mechanisms for planning various components of a sentence remain unclear. Eye-tracking studies targeting simple English transitive sentences (subject-verb-object) consistently find an “eye-voice span” of roughly 1 second—that is, a speaker’s gaze shifts to the picture of an object about 1 second before uttering its name (Griffin & Bock, 2000), suggesting a relatively narrow scope of planning. However, a recent series of studies suggests that advance planning is grammatically conditioned. For example, Momma et al. (Momma & Ferreira, 2019; Momma et al., 2016, 2018) showed that speakers plan verbs before the articulation of their grammatical object, but not before the articulation of their subject, suggesting that specific types of grammatical relationships among words determine aspects of their advance planning. Planning has been shown to be incremental in at least some cases—that is, the speaker may plan the

last parts of a sentence while uttering earlier parts, although such incrementality can be strongly influenced by strategic factors (e.g., Ferreira & Swets, 2002).

In general, it is accepted that a clause can be a unit of planning at some level of representation (Smith & Wheeldon, 1999). For instance, Ford and Holmes (1978) found that when English speakers were forced to respond to tones played in the midst of their five-minute extemporaneous monologues on a theme, their longest reaction times to the tones occurred near the end of a clause. Ford and Holmes interpreted these results to indicate that speakers conceive of their speech in one-clause units, and that planning for the upcoming clause occurs near the end of the current clause. Pawley and Snyder (2000) also concluded from an English corpus study that speakers plan one clause at a time, and a number of other studies have yielded results that imply clausal scope for planning (Beattie, 1980; Ford, 1982; Garrett, 1975; Meyer, 1996; Wijnen, 1990).

English and related languages that lack switch-reference marking have played a dominant role in the development of psycholinguistics (Mulak et al., 2021), so it is unsurprising that there is little in the literature to serve as a guide to how switch-reference in clause chains may be planned and produced. Smith and Wheeldon (1999) found that speakers took longer to begin coordinated two-clause sentences, such as [*The dog and the foot move up*] and [*the kite moves down*] than single-clause sentences, such as *The dog and the foot move up*. This was taken as an indication that some planning of the second clause already occurs before the speaker begins to utter the first clause. On the other hand, they also found that speakers were slower to start producing two-clause sentences that had complex first subjects (*the dog and the foot*), but simple second subjects (*the kite*) than two-clause sentences in which the first subject was simple and the second subject was complex. This was taken to show that speakers conceived of the second clause in a less detailed manner than the first clause during initial planning. Ferreira and Swets (2005) used pictures to elicit English sentences comprising three clauses: a main clause, an embedded subordinate clause, and another subordinate clause embedded within the first subordinate clause, such as [*This is the donkey that [doesn't know [where it is from]]*]. They showed that the amount of time that speakers took to begin the first clause varied depending on the grammaticality of the third clause, indicating that speakers were in some cases planning the entire structure in advance. These studies seem to support Garrett's (1982) proposition that sentence planning could sometimes span two clauses.

Clause chains are multiclausal sentences that differ from those tested by Smith and Wheeldon (1999) or Ferreira and Swets (2005). The simple coordinate structures tested by Smith and Wheeldon (1999) were conceptually repetitive, involving separate entities doing the same

action. In clause chains, consecutive clauses most often describe different actions. The structure targeted by Ferreira and Swets (2005) involved subordination, in which one or more clauses are embedded in a main clause; the clauses in clause chains are not embedded. Further, the first embedded clause in the Ferreira and Swets schema was a relative clause, and relative clauses (unlike clauses in clause chains) function to provide information about an entity that acts in the main clause. Given Ferreira's (1991) finding that relative clauses can be planned alongside the main-clause nouns they accompany, it could be argued that a relative clause functions as a part of the main clause rather than as a full additional clause.

Here we use a visual world paradigm to investigate comprehension and production of switch-reference marking in the Papuan language Nungon, spoken by about 1,000 people in remote villages of Papua New Guinea. To our knowledge, planning during sentence production has never been studied in a language with switch-reference marking.

In visual world eye-tracking, participants' gaze (on average) is assumed to reflect the focus of attention (Altmann, 2004; Altmann & Kamide, 2004, 2009; Huettig et al., 2011). Based on this working assumption, we can infer when participants begin processing a particular word or phrase by determining when their gaze shifts to the corresponding image. In Experiment 1, to understand whether listeners use Nungon switch-reference marking to predict during comprehension, we tracked participants' eyes as they were presented with audio recordings of brief narratives and images of characters in those narratives. In Experiment 2, to examine the time-course of Nungon speakers' planning of the subject of the upcoming clause, we tracked participants' looks to the current versus next subject while they recounted the same narratives.

Expanding research on cognition to communities outside industrialized societies brings challenges and compromises. For instance, while pressing keys on a laptop keyboard and answering multiple-choice questions are second nature to many reading this article, these are hardly natural in more remote communities around the world. Thus, among the challenges in field psychology is designing a task that is not so artificial that participants struggle to complete it, but not so open-ended that meaningful comparisons cannot be made. We therefore presented participants with naturalistic stimuli in the comprehension experiment and open-ended prompts in the production experiment. A feature of this design is that we were able to characterize processing that is more ecologically valid, although we lost some of the analytic power of comparing across controlled conditions.

These challenges are even more acute when experiments use advanced equipment—here, an eye-tracker. Running a portable eye-tracker that uses two laptop hosts at one time in a region without electricity was accomplished through a

long-term solar system setup in the Nungon-speaking area, with enough power to run both display and control laptops.

Strong community relations are crucial to the success of field-based experiments, and to laying the foundations for further work with the same community. If community members are uncertain about the intentions of a researcher, or the purpose of the research, they may abstain from participation and decide not to support similar research in the future. The first author has maintained a close relationship with the Nungon-speaking community of Towet village since 2011, beginning with immersion linguistic fieldwork there. She is adopted into a local clan.

Months before the research team traveled to Towet to run the suite of experiments that included these eye-tracking experiments, Towet community members Stanly Girip, James Jio, and Lyn Ögate began planning for the “experiment fair” of which the current experiments were a part (see Method). They recruited four research assistants from among Towet adults who had obtained at least a 10th-grade diploma (a rare accomplishment, requiring boarding at distant schools), and convinced all 30 households in Towet village to take two weeks off from all regular duties in order to be available as participants for the planned experiments. This 2-week break from farming was possible because the community stockpiled crops and firewood for months to ensure that no one would go hungry. Overall, the Towet village community went to extraordinary lengths to ensure the success of these experiments. Their major effort is testament to the specialness of this community, and to the first author’s long-standing collaborations with them (see Dobrin, 2008, on the importance of long-term research collaborations in Melanesia).

Switch-reference marking in Nungon

Nungon is a Papuan language of the Finisterre-Huon family, spoken in six villages in the Uruwa River valley in the Saruwaged Mountains of Morobe Province, Papua New Guinea (Sarvasy, 2017). There are about 1,000 speakers, but—typifying the staggering diversification of languages in Papua New Guinea—they are spread across six distinct dialects, with no more than about 350 speakers of any one dialect. All local people grow up with Nungon as their first language; most have some familiarity with the English-based creole Tok Pisin, but this is not used outside the local schools and church services. Basic literacy levels in Nungon and Tok Pisin are high, but most adults do not read or write on a daily basis. The Uruwa River valley is remote and accessible only by small plane or foot (a difficult multi-day hike through alpine forests to the port city of Lae). The region lacks electricity and only recently gained a cell phone tower. Most adults work as self-sufficient small-holder farmers.

The community is special, even in an overwhelmingly rural nation like Papua New Guinea, in that they rejected the notion of establishing an internal market economy, in favor of maintaining age-old traditions of sharing crop surpluses.

The Nungon language has complex verbal morphology. For instance, verbs can be marked for one of five tenses. Subject and object noun phrases are often omitted in Nungon discourse (“*argument dropping*”). Like English, Nungon has clausal coordination and subordination. However, clause chains are extremely common; for instance, text messages in Nungon often comprise one or more clause chains with four or more clauses apiece (Sarvasy, 2021). Clause chains are highly predictably distributed in narratives, but other sentence types, which lack switch-reference marking, can predominate in other genres. In a sample of 49 Nungon narrative monologues (including 1,742 clause chains), the longest clause chain had 22 clauses, while the average length was 3.4 clauses (Sarvasy, 2021).

The verbs in nonfinal clauses in a Nungon clause chain are obligatorily marked with a switch-reference suffix. These suffixes encode two different possibilities: same-subject (SS), after which the subject of Clause A is maintained in the following Clause, B, and different-subject (DS), after which the subject of Clause B differs from that of Clause A.

3. [Kurawiöng o-unya],
Kurawiöng descend-DS.2/3DU

[urop y-aa-gu-ng], amna nangnang.
enough 3NSG-SEE-REMOTE.PAST-2/3PL¹ man eater

“The two of them descending at Kurawiöng-SWITCH, that’s it, they saw them, man-eaters.”

Example (3), from a recorded narrative (and one of the audio stimuli for Experiment 1 here), illustrates the Nungon penchant for omission of subject and object arguments. The first clause has just a single proper noun (a place name), followed by a DS-marked verb. In Nungon, the DS suffixes encode both DS marking and the person/number of the current clause’s subject, while the SS suffix encodes only SS marking, and involves no subject agreement. In the first clause here, the DS suffix is the only grammatical indication that the subject is second or third person and dual number (that is, two): there is no subject noun phrase in the clause. The second clause has an adverb and a verb that is inflected for remote past tense, and both subject and object person/number; again, subject and object are referenced solely through affixes on the verb, which is always

¹ Abbreviations: DS (different-subject), 2/3DU (second or third person, dual number), 3NSG (third person, non-singular), 2/3PL (second or third person, plural number); dual = exactly two; non-singular = more than one; plural = more than two.

the final element in the clause. As it happens, the object of the second clause refers to the pair of man-eaters who are the implied subject of the first clause. An explanatory noun phrase, “man-eaters,” follows the second clause.

Since this sentence occurs in the middle of a narrative, characters and situation are understood from the established discourse context. In such a small, close-knit community, omission of subject and object arguments in quotidian conversation more generally is enabled by the fact that people often share much background information about events and people in their communities (cf. Wray & Grace, 2007).

Nungon switch-reference strictly tracks grammatical subjects, even when someone other than the subject is the real actor. For instance, in expressions like “I feel angry,” Nungon speakers actually put “anger” as the subject of the verb, and “me” as the object: *iik na-mo-ha-k* “anger 1SG-give-PRESENT-3SG,” or, roughly: “anger affects me.” Several negative emotions and sensations are described in this way, such as “feeling tired,” “feeling heavy,” and “feeling bored.” Crucially, because Nungon switch-reference marking strictly tracks the syntactic subject, even when the “notional” subject does not change from clause to clause, speakers use DS markers prior to expressions like these. For instance, in (4), even though the notional subject remains the same throughout, the syntactic subject changes from “I” to “anger” to “I” again, so a speaker must use the DS marker at the end of each nonfinal clause:

4. *[E-waya], [iik na-m-una],*
 come-DS.1SG anger 1SG-give-DS.3SG

[bög-in ongo-go-t].
 house-LOCATIVE go-REMOTE.PAST-1SG
 “I coming-SWITCH, anger affecting-SWITCH me, I went home.”

This implies that there must be a detailed grammatical element to switch-reference planning, such that it does not just occur at a broader conceptual level.

Children learning Nungon produce two-clause chains by age 2.5, and three-to-five-clause chains beginning around age 3 (Sarvasy, 2019, 2020). Both SS and DS markers are evident in their early clause chains, and 60%–80% of switch-reference morphemes in parental speech are SS.

Experiment 1: Comprehension

An intriguing possibility is that switch-reference marking could exist in part to provide comprehenders with a cue that might facilitate processing of the subsequent clause. This may be especially helpful in an argument-dropping language like Nungon, where subjects are sometimes not overtly expressed. To understand how switch-reference



Fig. 1 Research assistant Lyn Ögate runs a participant through the comprehension experiment, Towet village

marking affects online processing during comprehension, we tracked participants’ gaze while they listened to 15 short speech samples that included clause chains (Fig. 1). We expected that comprehenders’ fixations would differ depending on whether they heard an SS or DS switch-reference marker. The precise timing of this difference would enable us to assess whether speakers use the morphemes as cues for predictive processing. We expected that, in a DS condition, comprehenders’ gaze would begin to shift away from the “same subject,” or the subject of the clause the switch-reference marker appears in, before the identity of the next subject was clarified in the next clause.

Method

Participants

Sixty-six adult participants were recruited from Towet village, Uruwa Ward 1, Kabwum District, Morobe Province, Papua New Guinea. Participants were each paid 50 Papua New Guinean kina, approximately 15 U.S. dollars.² Participation occurred as part of a four-experiment “science fair” (see Mulak et al., 2021). Local project managers oversaw recruitment. Participants were read an information sheet in Nungon before starting the experiment, and signed a consent form. Two participants were later excluded because they were nonnative Nungon speakers who had married into the region from elsewhere; all other participants were native speakers of Nungon. Three other participants’ data were not

² Since people in the region do not pay each other for labor, nor do they buy or sell staple foods, it is hard to quantify this amount in terms of a local standard, but for comparison, the national minimum wage is 3.50 kina per hour.

IP Time: 0000000 ms / Trial Time: 00003034 ms



Fig. 2 A screenshot of the experimental display, showing interest areas in boxes. This screen accompanied the stimulus sentence [umar-a], worok, [ingguk umar-uya], [urop, amna temogok]. ‘[Dis-

mantling-NOSWITCH], thus, [(they) dismantling-SWITCH one], [that’s it, (he) shot a man].’

recorded by the experimental software, such that 61 participants’ data were included in analyses.

Materials

From a corpus of more than 200 Nungon personal experience narratives compiled during fieldwork on the Nungon language (Sarvasy, 2017), 15 short audio stimuli were selected. These stimuli ranged in duration from 2 to 29 seconds (mean duration: 9.7 seconds, standard deviation: 7.1 seconds), and had been recorded by nine different adult speakers (five males). Stimuli were selected if they comprised at least one clause chain, including at least one switch-reference marker; were easy to visually represent; and were produced clearly. Eight stimuli involved two different, nonoverlapping grammatical subjects, four stimuli involved three different grammatical subjects, one stimulus involved five different grammatical subjects, and two stimuli involved only a single grammatical subject. These two stimuli were also the only stimuli to lack DS switch-reference markers altogether. In the other 13 stimuli, either the sole switch-reference marker was DS, or of multiple switch-reference markers, one or more were DS.

Each audio stimulus was paired with a display comprising one interest area for each subject argument in the audio stimulus. This meant that: the displays for the eight audio stimuli with two grammatical subjects had two interest areas, placed apart on the screen, either in different corners or far apart along a horizontal axis; the displays for the four stimuli with three grammatical subjects had three interest areas,

again, spread apart on the screen, and the display for the stimulus with five grammatical subjects had five dispersed interest areas. In the displays for the two stimuli with just a single subject maintained throughout the clause chain (and only SS switch-reference markers), there were two interest areas: one containing a representation of the actual subject, and another containing a “distractor” image. Displays were hand-drawn by the first author; characters depicted wore culturally appropriate clothing and used appropriate tools (such as bows and arrows and string bags, as mentioned in the stimuli). An example of a display is in Fig. 2; here, interest areas as programmed into the experiment are shown with boxes; the pink circle shows gaze at one time-point within the lower interest area. Note that the two subjects in the stimulus accompanying the display in Fig. 2 are ‘they’ and ‘he.’ Looks to the individual men within the upper-left-hand interest area were not differentiated for the purposes of the experiment, and this is the case for all dual and plural subjects.

Procedure

The experiments were run in one room on the second floor of a purpose-built building with woven bamboo walls and floors in Towet village, in the Nungon-speaking area. The building is equipped with three 100-W solar panels and accompanying 12-V batteries, charge controllers and AC/DC inverters. The eye-tracking experiments were part of an “experiment fair,” in which four foreign researchers, four local research assistants, and three local organizers ran four psychological and psycholinguistic experiments over 2 weeks in mid-2019.

Each experiment took place in one room of the building or in a temporary enclosure outside. Local organizers tracked community members' participation in the five experiments, such that participants moved seamlessly between experiments, and all those who wished to participate in all four experiments could do so (see also Mulak et al., 2021). The eye-tracking experiments were run jointly by the first author and organizer Lyn Ögate, who took turns running participants.³

The eye-tracking comprehension and production experiments were created as a single experiment using Experiment Builder software (SR Research) and administered using an EyeLink Portable Duo eye-tracker, which recorded participants' eye movements while they listened to and produced sentences. Participants were seated a comfortable distance from the presentation laptop and a target sticker was placed on each participant's forehead. This allowed accurate eye-tracking without impairing movement (e.g., during production). Viewing was binocular, but fixation location was monitored from their right eye following a 9-point calibration.

Participants were tested in one session lasting approximately 30 minutes, with the experiment divided into two blocks—comprehension and production. All participants first completed the comprehension block before the production block, though items were randomized within blocks for each participant.

Before the comprehension block, participants were told that they would need to keep their eyes on the screen while listening to speech in Nungon. Each trial began with the presentation of a fixation cross in the center of the screen. To control looking bias, the experiment was programmed so that the visual scene appeared only after participants had fixated on the cross for 500 ms. Then, 1,000 ms after the scene was presented, an auditory stimulus sentence was played over headphones. The experimenter pressed the space bar to move onto the next item once the recording was finished. Each recording was presented once.

Unfortunately, during testing, the experiment software repeatedly crashed, which could only be worked around by using a “demo” version of the experiment. This version was identical to the licensed version, except that it had the words “DEMO VERSION” in approximately 12-point red type in the center of each display screen. We saw no evidence during experimentation that participants' eyes were drawn to these words. In the end, 25 participants of the original 66 completed the experiment using the demo version of the display.

Analysis

Using Praat software (Boersma & Weenink, 2019), switch-reference markers in the 15 audio stimuli were coded, and their onset and offset times extracted. Where a

switch-reference marker occurred on a verb that also bore an object prefix referring to a character in another interest area, this was excluded from consideration here. Switch-reference markers in clauses with unclear or ambiguous reference were also excluded from the analysis. Where preceding material undergoes phonological change with the addition of a switch-reference suffix, the onset of the syllable before the suffix was extracted; otherwise, the onset of the morpheme itself was extracted. These were then coded for the subject of the clause in which they occurred. Finally, the onset of the morpheme's own clause and the onset of the following clause were extracted.

Prior to analysis, eye-tracking data were epoched into 1,500 ms trials, each time-locked to the onset of the switch-reference morpheme. The 15 stimuli combined included a total of 49 switch-reference markers (23 DS and 26 SS). Each of these morphemes was treated as an independent stimulus. The model was thus fed data from 49 items per participant.

Since each interest area was the visual representation of a grammatical subject, the eye-tracking data could be analyzed in terms of whether, for each trial, a participant was looking at the subject of the clause bearing the switch-reference morpheme, or not. In other words, we investigated gaze patterns after the onset of the switch-reference marker in terms of whether, at each time point, the participant looked to the interest area depicting the subject of the first clause (“looks to same subject”). Note that for stimuli including more than one switch-reference marker and at least one DS marker, the interest area associated with “same subject” can change for each trial (each switch-reference marker within the stimulus). In the English pseudo-clause chain in (1), for instance, the “same subject” for “finding its bowl empty” would be “cat,” while the “same subject” for “the dog chasing it” would be “dog.” This means that all data pertaining to one stimulus could not simply be coded according to “looks to interest area A.”

Eye-tracking data were thus coded in a binary fashion as “looking at same subject”—the subject of the clause with the switch-reference morpheme—or “looking at another interest area.” Time points were excluded if the participant was not looking in one of the pre-defined interest areas on the screen (or when the participant's eye could not be detected by the eye-tracker).

The data were analyzed using a logistic mixed effects regression (Baayen et al., 2008). We analyzed two factors: *morpheme type* had two levels, SS (same subject) and DS (different subject), and was treatment-coded. Because we expected any effect of *morpheme type* to emerge over time, we included *time* as a continuous factor. Gaze data were sampled in 150 ms intervals starting at morpheme onset (time zero) and ending 1,500 ms later. Prior to analysis, the

³ While not a native speaker, the first author has been traveling to the area since 2011, has been adopted into the local kinship system, and is fluent in Nungon.

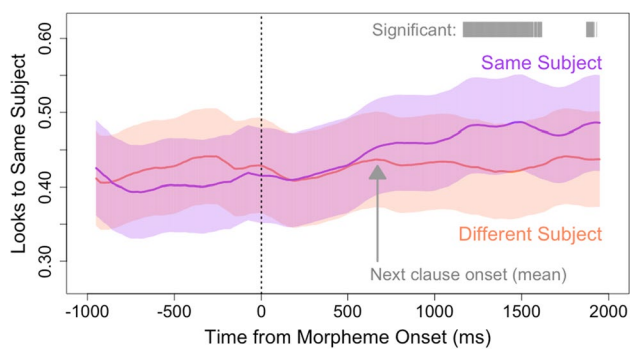


Fig. 3 Looks to the same subject (i.e., the subject of the current clause) by condition (smoothed with 100-ms rolling average for visual clarity; all analyses performed on unsmoothed data)

time variable was centered and scaled such that it ranged from -1 to 1 .

Following Barr et al. (2013), we report the results of the model with the maximal random effects structure that converged, having removed random effects in order of least variance accounted for to most. In addition to fixed effects terms for *morpheme type*, *time*, and their interaction, the final model had random intercepts for participants and items, and random slopes for *morpheme type* within participants.

Results

Results are shown in Fig. 3. At morpheme onset ($t = 0$), the proportion of looks to the same subject was roughly equal in the two conditions. In the DS condition, the frequency of looking to the same subject was relatively constant over time. But in the SS condition, looks to the same subject increased with time, leading to a significant difference starting 1,164 ms after morpheme onset (grey bar). This was after the mean onset time of the next clause (672 ms after morpheme onset; arrow).

The model detected no differences in looks to the same subject between DS and SS conditions when collapsing across time (the main effect of *morpheme type* was not significant, $\beta = -0.504$, $z = -0.690$, $p = .490$). The model also failed to detect a significant change in the proportion of looks to the same subject over time when collapsing across DS and SS conditions (the main effect of *time* was not significant, $\beta = 0.011$, $z = 0.333$, $p = .739$). Crucially, however, the model did detect an increasing tendency over time to look at the first subject in the SS condition relative to the DS condition (the interaction of *morpheme type* and *time* was significant, $\beta = 0.256$, $z = 5.608$, $p < .001$).

To determine the earliest point at which there was evidence for a difference between the SS and DS conditions, a series of 1000 fixed-effects-only logistic regressions analyzing looks as a function of *morpheme type* were performed

on each sample between 0 and 2,000 ms (i.e., one model for each sample at a sampling rate of 500 Hz). The resulting 1,000 p values for the *morpheme type* term were FDR-corrected for multiple comparisons. Time points for which these adjusted p values were below .05 are indicated with the grey bar in Fig. 3. The earliest time that showed a significant difference between the SS and DS conditions was 1,164 ms after morpheme onset.

Discussion

The results of the comprehension experiment failed to support the notion that listeners use switch-reference markers as cues for prediction. As expected, participants looked more to the same subject after hearing a SS morpheme than after a DS morpheme. However, timing indicates that this difference does not stem from information in the morphemes themselves. While the divergence in looks between the two conditions appears to begin around 500 ms after morpheme onset, the difference does not achieve significance until 1,164 ms after morpheme onset. This is well after the mean onset time for the next clause, which was 672 ms after morpheme onset, indicating that the difference likely does not reflect predictive processing. Indeed, by the time the difference achieves significance, participants have in most cases had several hundred milliseconds to discern the identity of the next subject from information contained in the next clause.

Why do listeners not seem to look immediately away from the current subject, on hearing a DS morpheme? If valid, our results here could indicate that the information carried by the switch-reference morpheme is not used during comprehension, although previous findings about the use of morphological cues to guide comprehension in European languages (Hanne et al., 2015; Meir et al., 2020) and non-European languages (Mitsugi, 2017), imply that this is improbable and that the switch-reference morpheme should help guide comprehension at some level.

Alternatively, the key could lie in the amount of information encoded in the morphemes themselves, together with the nature of the visual world task. DS marking in Nungon, as in most languages with switch-reference,⁴ simply indicates that the upcoming subject will differ. It does not necessarily help the listener determine who or what the new subject will be. In an artificial task with just two visual interest areas to choose from (say, A or B), each representing an actor, a listener might be able to

⁴ In fact, a few languages, such as the Papuan language Hua (Haiman, 1980), exhibit “anticipatory” switch-reference marking, in which both the person/number of the current subject and that of the upcoming subject are specified on nonfinal verbs in clause chains.

use switch-reference morphemes to guide prediction (since *not-A* could imply B). But in a task with more interest areas, and, for that matter, in natural discourse, where the choice of upcoming subjects is unconstrained, the information encoded through switch-reference morphemes is insufficient to choose an alternative interest area/upcoming subject (since *not-A* could imply B or C). In such situations, the listener may well wait until more information is available to actually shift gaze from the current subject.

To test this possibility, we re-ran analyses using only the set of data from the 10 stimuli with two visual interest areas, but no clearer picture of use of switch-reference morphemes emerged from this modeling. This does not necessarily mean that the account outlined above is incorrect: Participants could remain open to the possibility that an upcoming clause could have a subject not depicted on the screen, in which case the number of images should not necessarily be expected to constrain predictions. Further, although these 10 stimuli have only two interest areas each, listeners could still attend to switch-reference morphemes in their usual way, which could be, in discourse, to wait for clarification in the upcoming clause itself before predicting its subject. Finally, it is possible that participants do in fact use switch-reference morphemes for prediction, but that this is not reflected in patterns of looking.

Of course, these results could also be clouded by the inherent problem of using naturalistic stimuli: While they afford ecological validity, because conditions are not controlled manipulations, there is no guarantee that other cues and processes were the same in the two conditions. Had all else been kept equal, it is possible that a difference in gaze would have emerged much earlier.

An ideal follow-up to the present work, then, would be to attempt to replicate the findings with controlled stimuli. Specifically, recordings could be spliced such that narratives come in pairs of stimuli which are identical up until the verb, at which point a verb with an SS morpheme is spliced into one of the recordings and a verb with a DS morpheme spliced into the other. A number of other considerations would likely be important to control, such as the a priori likelihood of an SS versus DS morpheme at that point in the stimulus, as well as the number of candidates for the next subject on the display. Such a design would allow us to disregard the possibility that any differences observed (or not observed, as in the present study) are due to differences in the preceding context, and to directly interpret any differences as reflecting switch-reference morpheme-specific processing.

Although the DS morpheme can be analyzed as providing insufficient information for a listener to fully predict the identity of the upcoming subject, the situation is different for a speaker. The speaker is obligated to produce a switch-reference morpheme, and it seems that they must process the

upcoming clause's subject in order to produce the correct morpheme. We investigate this possibility in Experiment 2.

Experiment 2: Production

In production, we aimed to estimate how far in advance (in seconds, and in clauses) speakers plan when they utter clause chains in Nungon, in the hopes of comparing this to estimates based on processing of more heavily studied languages like English, German, and Japanese. We did so by presenting the same group of participants in the comprehension experiment with the same images viewed during the comprehension experiment, but asking participants to narrate the story for each set of images themselves. We then determined when looks to the same subject diverge in the seconds leading up to production of either an SS or a DS switch-reference marker. An estimate of about 1 second, or one-to-two clauses, of advance planning would be consistent with previous experimental literature, and would validate this finding with data from a vastly different language and population. Another possibility would be that, because the syntax of Nungon requires advance planning of the next clause in a way that more heavily studied languages' grammars do not, Nungon speakers would plan even farther in advance. This would call into question the generality of previous estimates of the scope of advance planning, and would highlight the need for psycholinguistic research on a more diverse set of languages and participant populations.

Method

Participants

The same 66 Nungon-speaking adult participants completed the production experiment immediately after the comprehension experiment. This was treated as part of the same task, so there was no separate consent process, nor a separate payment. As described above, the original 66 participants were winnowed to a final group of 61, due to loss of data and non-native speaker status. Of the 61 participants whose data were included in the comprehension experiment analyses, however, two had no productions that included clause chains, so the final participant pool for the production experiment includes 59 adults.

Materials

Stimuli consisted of the same 15 displays presented during the 15 audio stimuli in the comprehension experiment. Images in interest areas appeared in the same locations in each display as in the comprehension experiment.



Fig. 4 Image corresponding to comprehension stimulus in (3)

Procedure

Prior to the start of the production experiment, participants' eyes were recalibrated. They were told that they would see the same displays they had seen before (in the comprehension experiment), but that this time they themselves were to tell the story of the characters in the display. They were free to either retell what they had heard before, or, if they did not remember that story or preferred not to retell it, they could tell a different story, or simply describe the scene that they saw. Each of the same displays from the comprehension experiment was then shown, in random order. After each display appeared, the experimenter would ask the participant in Nungon if they were ready, and then the experimenter controlled the beginning and end of the recordings by pressing the space bar on the display laptop.

Analysis

Each participant recorded 15 short stories or descriptions, producing a number of switch-reference morphemes in the process. Figure 4 shows the display that was shown along with the stimulus given in (3) above during the comprehension experiment. There are two interest areas here, one in the upper left, containing an older couple, and one in the lower right, containing a group of children. The translations of five sample Nungon productions related to the image are in (5).

5. Sample English translations of Nungon productions based on image in Fig. 4 (clause boundaries marked with brackets):

- a. [*A man with a woman being-SWITCH like that*], then, [*three children were all there*].
- b. [*Two people coming-SWITCH*], [*(the others) seeing-NOSWITCH them*], then, [*being-NOSWITCH afraid of them*], [*(they) were preparing to go*].
- c. [*Girls, coming-SWITCH with their mother*], [*a man, coming-NOSWITCH with a woman*], [*were on the side*].
- d. [*Wearing-NOSWITCH a grass skirt and such*], [*(they) were relaxing*].
- e. [*The two of them were descending from a ridge*]. [*A woman and children being-NOSWITCH at home*], [*were waiting*].
- f. [*(As) little girls were staying-SWITCH there*], [*a demon couple coming-NOSWITCH*], [*chasing-NOSWITCH them*], [*coming-SWITCH*], [*(the girls) coming-NOSWITCH*], [*were looking at them*].

All productions by all 61 native speaker participants for whom data were recorded were analyzed by hand. Using Praat, the onset and offset of each switch-reference morpheme for which the subject was clearly identifiable, and where the host verb was not also marked with a potential competitor for gaze (e.g., an object prefix referring to the characters in another interest area), were extracted. Then, as

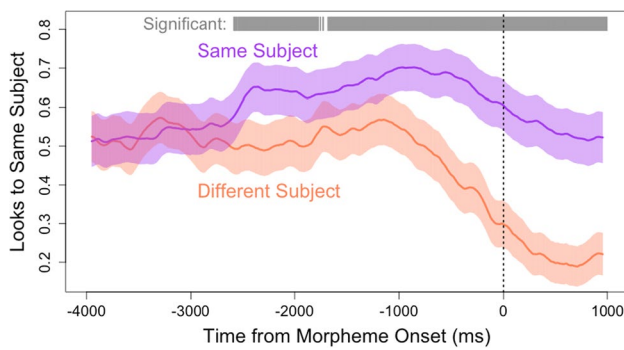


Fig. 5 Experiment 2 results: Looks to the same subject as a function of time benchmarked to the onset of the switch-reference morpheme ($t = 0$ on the horizontal axis), separately for same subject and different subject switch-reference morphemes. Grey bar at the top of the figure indicates a significant difference between morpheme conditions

with the switch-reference morphemes in the comprehension stimuli, each switch-reference morpheme was also coded for the subject of its clause. This resulted in a total of 768 eligible switch-reference morphemes (350 DS and 418 SS). Each of these morphemes was regarded as an individual “trial.” Eye-tracking data were epoched into 1,500 ms windows, each beginning 3,000 ms prior to the onset of a switch-reference morpheme.

The modeling approach was exactly the same as for the comprehension experiment. The maximal logistic mixed-effects regression had fixed effects for *morpheme type* and *time*, which was again sampled in 150 ms intervals between 3,000 to 1,500. The model included a random intercept for participants, within which both *morpheme type* and *time* were allowed to vary.

Results

Results of the production data analyses are in Fig. 5.

Figure 5 shows that when speakers produced SS morphemes, looks to the same subject steadily increased from around 3,000 ms prior to morpheme onset. When producing DS morphemes, the rate of looking at the same subject remained relatively steady until around 1,100 ms pre-onset, at which point looks to the same subject decreased dramatically.

The model detected that in the analysis window (i.e., $-3,000$ to $-1,500$ ms relative to morpheme onset), participants looked more to the same subject in the SS condition than in the DS condition, collapsing across time (the effect of *morpheme type* was significant, $\beta = 0.553$, $z = 2.927$, $p = .003$). The model did not detect any change in looks to the same subject over time when collapsing across DS and SS conditions (the effect of *time* was not significant, $\beta = 0.059$, $z = 0.088$, $p = .500$). The model found no evidence for a difference between DS and SS conditions in the change in

looking patterns over time during the $-3,000$ to $-1,500$ ms analysis window (the interaction between *morpheme type* and *time* was also not significant, $\beta = .135$, $z = 1.492$, $p = .136$).

As in the comprehension experiment, a series of 2,500 fixed-effects-only logistic regressions were run spanning the window from $-4,000$ ms to $1,000$ ms preonset (i.e., one model every 2 ms). FDR corrections were performed on the p values for *morpheme type* estimates (see grey bar in Fig. 5). The earliest significant difference between looks in the DS and SS conditions occurred at $-2,586$ ms.

Discussion

As predicted, Nungon speakers seem to plan subjects, at least at some level of representation, in clause chains long before articulating the switch-reference morphemes that precede them. In fact, our results show differences in gaze depending on whether the speaker produces an SS or DS morpheme over 2.5 s before the onset of the morpheme.

Previous research involving English-speaking participants has shown that (a) the eye-voice span is generally 1 s (Griffin & Bock, 2000), and (b) speakers are capable of planning one clause in advance (references in Smith & Wheeldon, 1999), and possibly up to two clauses in advance under certain conditions (Ferreira & Swets, 2005; Garrett, 1975; Smith & Wheeldon, 1999), although more incremental planning appears to be possible depending on the particular task demands (Ferreira & Swets, 2002; Wagner et al., 2010).

In examining gaze relative to uncontrolled utterances, we have expanded on the methods of Smith and Wheeldon (1999) and Ferreira and Swets (2005), who each studied one very specific structure (e.g., for the former paper: coordination of two clauses with a single verb). Our results could be interpreted as indicating that Nungon speakers plan switch-reference morphemes over two times farther in advance than the typical eye-voice span estimate predicts.

We can estimate the number of clauses in advance that this advance gaze shift represents. Based on a representative sample of 104 clauses from clause chains in the production experiment, we calculated that the average clause duration was 1,536 ms (standard deviation: 960 ms). To estimate how many clauses in advance Nungon speakers plan before starting to produce a clause, we compared this number to what we refer to as the *planning span*: an estimate of the amount of time before the next subject that looks diverged in the DS and SS conditions. The planning span was estimated by adding the amount of time pre-morpheme onset for which there was evidence of a difference between conditions (2,586 ms) to the average duration of the rest of the clause (377 ms), giving a total duration of 2,963 ms. Note that this is a conservative estimate, in that it ends at the offset of the current clause rather than the onset of the next clause, which in

many cases began after a brief pause. We divided the planning span (2,963 ms) by the average clause duration (1,536 ms), which gave an estimate of 1.929 clauses in advance, on average.

The implications of this span depend on what exactly is being planned. Up to here, we have talked about planning as if there were some scope within which every feature of an upcoming utterance is spelled out in perfect detail. But this is merely a helpful simplification. In reality, the degree of advance planning varies for different levels of representation (e.g., morphosyntactic vs. conceptual).

Although it is not possible to definitively tell which levels of representations speakers planned the next subject in advance in the current experiment, it is likely that this advance planning occurs at the grammatical, not just conceptual, level. This is because, as explained above, knowledge of the conceptual role (like the agent) is insufficient to determine which switch-reference marker to use. Speakers need to decide which event participant will be grammatical subject in the upcoming clause to be able to decide which switch-reference marker to apply in the current clause. Based on this reasoning, the divergence in gaze that we detect beginning around 2,586 ms prior to morpheme onset reflects may suggest that speakers in our study grammatically planned the subject on the order of 3 s in advance (i.e., the planning span calculated above). However, it remains unclear if speakers planned the lexical representations of the upcoming subjects, especially when the upcoming subject is different from the current subject. When the upcoming subject is different, speakers may postpone the lexical retrieval processes associated with the upcoming subject, because they can in principle decide use a switch-reference marker without necessarily identifying which word to use for the next subject. That is, speakers may first plan whether the next subject is the same or different, before planning which lexical item to use if the next subject is planned to be different. Indeed, an anonymous reviewer pointed out that there seem to be two stages of divergence in the gaze data: first, an increase in looks to the same subject in the SS condition around $-2,500$ ms, and then a steep drop-off in the DS condition around $-1,000$ ms. While the uncontrolled nature of our stimuli makes it hard to know for sure what drives these two aspects of the data, one reasonable hypothesis is that the early divergence reflects a difference in syntactic planning, and the latter reflects more fine-grained lexical planning.

In either case, our findings suggest that syntactic planning can happen earlier than previously known. Garrett (1975), Ford and Holmes (1978), Smith and Wheeldon (1999), and Ferreira and Swets (2005) showed that syntactic (not semantic) planning occurs at the scale of the single clause. Our data would seem to indicate that syntactic planning, at least under certain circumstances in Nungon, can happen almost two clauses in advance.

General discussion

This paper presented two psycholinguistic experiments investigating language processing in Nungon, an understudied Papuan language spoken by about 1,000 people in remote villages in Papua New Guinea. We investigated how speakers comprehend and produce sentences containing switch-reference morphemes, a form of cross-clause verb agreement that is unlike any we are aware of in more commonly-studied languages. Switch-reference marking requires speakers to inflect each nonfinal verb in a clause chain with a suffix that specifies whether the subject of the next clause will be the same or different. Thus, Nungon morphosyntax requires speakers to have planned at least the subject of the next clause in a chain prior to completing the current one. This raises the intriguing possibility that Nungon speakers may plan farther in advance during sentence production than speakers of languages like English. This, in turn, would call into question the generality of previous work—based largely on English—which aimed to determine how far in advance speakers plan their utterances.

In Experiment 1, a visual world comprehension study, participants listened to naturalistic recordings of short narratives while they viewed images of the characters involved in the narratives. Listeners did not appear to use the switch-reference morphemes as cues to prediction. One explanation centers around the limited information provided by the DS marker about the identity of the upcoming subject: It could be that listeners must complement switch-reference markers with other cues to form full concepts of the next clause's subject.

In Experiment 2, a visual world production study, participants from Experiment 1 were shown the same sets of images they had seen in the comprehension experiment, and were asked to reproduce the narratives. We measured how far in advance speakers' gazes reflected whether they produced an SS or DS morpheme, and estimated that this occurs roughly 1.9 clauses prior to the onset of the next clause, on average. It is widely accepted that speakers of English and some other languages can plan speech up to one clause in advance, and some studies have found evidence for planning of at least a portion of a second clause in advance. That said, the studies that found indications of two-clause advance planning at some level did so using a very limited set of specific sentence templates (Ferreira & Swets, 2005; Smith & Wheeldon, 1999), while we have now shown this using much more varied, naturalistic data.

Cognitive science studies sometimes employ the assumption that what is true of samples of university students in industrialized countries is true of humans in general (see Huettig, 2015, for discussion of this assumption).

Language research in particular is susceptible to potential bias, as many of the languages on which theories of linguistic structure and processing are based are related to one another, and therefore have similar properties. For instance, it is quite common for verbs to agree with their own subjects in European languages, and this has given rise to an entire literature on how agreement is tracked and processed. However, even among psycholinguists, few know that there are languages like Nungon in which verbs reflect not only their own subject, but also the one that comes next.

Limitations and future directions

As mentioned above, one of the difficulties in running experiments with Nungon speakers is the population's lack of exposure to these kinds of tasks. We therefore used naturalistic stimuli in the comprehension experiment, and relatively open-ended prompts in the production experiment. While this comes with the benefit of higher ecological validity, it also means that directly comparing behavior across conditions must be qualified. For instance, we argued that the lack of evidence for predictive processing in Experiment 1 might be a result of other differences between the stimuli in the two conditions. Similarly, differences between conditions such as those observed in both experiments could also conceivably be the result of extraneous differences.

In Experiment 2, when participants tried to re-tell what they had heard before, it is possible that how they planned sentences differed from spontaneous production. However, many previous psycholinguistic studies suggest that various production effects, including the syntactic priming effect (Lombardi & Potter, 1992), the availability effect on word order (McDonald et al., 1993), and the semantic interference effect and its timing in sentence-level production (Momma & Yoshida, 2021) can be observed in recall tasks, just like in picture-based tasks. As those effects are usually interpreted to arise from the dynamics of sentence planning (Bock & Ferreira, 2014; Chang et al., 2006; Momma & Yoshida, 2021), we believe that the conclusions about the temporal properties of sentence planning in the current study would not be invalidated even if a majority of our speakers indeed tried to retell the story rather than produce novel utterances spontaneously. In any case, one important future direction will be to attempt to validate the current findings with more controlled stimuli.

We also aspire to investigate the limits of advance planning of Nungon clause chains. The finding of one-to-two-clause advance planning does not necessarily mean that speakers cannot plan farther ahead. Indeed, switch-reference marking may provide an opportunity to probe just how far planning can go. Throughout this paper, we have assumed that switch-reference marking only entails planning the

subject of the next clause. However, Momma and colleagues (Momma & Ferreira, 2019; Momma et al., 2018) suggested that speakers may use verbs' lexical information to grammatically encode the subject specifically, when the verb is of a type known as *unaccusative* (e.g., *boil, fall, grow*). If the present production experiment were repeated with unaccusative verbs, one might be able to detect advance planning beyond the span we posit here.

We hope that the present work can serve as a model for future collaborations between field linguists and psycholinguists. It is crucial that psycholinguistics validate findings with data from understudied languages and nonindustrialized, nonuniversity-educated participants. Conversely, unless field linguists can connect their research on out-of-the-way languages to broader issues in the cognitive sciences, the astounding structural singularities of these languages will languish in obscurity, unrecognized beyond specialist ranks (Evans & Levinson, 2009).

An important component to this study is to deliver the results to the Nungon-speaking community, who were pleased to begin to think about how clause chains differed from English-style sentences in an initial "grammar workshop" run at the community's behest in 2017. Like many small speech communities throughout Papua New Guinea and the world, they teeter on the brink of language shift: just a few hundred migrants to the urban fringe away from losing their heritage. Learning about why complex features of their language hold interest for cognitive scientists around the world would help enable them to make an informed decision about maintaining their language and culture in the face of global changes.

Conclusion

We have presented two experiments that aimed at understanding language processing in Nungon, an understudied language of Papua New Guinea. Nungon switch-reference marking, a type of cross-clause agreement absent from more commonly studied languages like English or German, requires speakers to indicate in advance whether the subject of the upcoming clause will be the same as, or different from, that of the current clause. The comprehension experiment gave inconclusive evidence for whether listeners use switch-reference markers as prediction cues. The production experiment gave strong evidence for advance planning of switch-reference morphemes during naturalistic speech production, beginning about two clauses in advance, although exactly what type of representation is planned at this early stage remains to be determined. Overall, these results demonstrate that Nungon speakers are able to plan multi-clause sentences roughly two clauses in advance, at a morphosyntactic level, with triple the eye-voice span previously attested for English.

Acknowledgments We would like to thank the Nungon community for their major efforts in supporting this work and participating in the experiments. Nungon speakers Stanly Girip, James Jio, and Lyn Ögate ensured that the experiments ran smoothly, and Lyn Ögate ran the experiment, in some instances, independently. Laurence Bruggeman and the MARCS Institute Technical Team provided vital technical assistance before the field trip, and Alba Tuninetti helped troubleshoot during the field trip. Helpful feedback at various stages was received from Alexandra Aikhenvald, Laurence Bruggeman, Anne Cutler, Paola Escudero, Nicholas Evans, Evan Kidd, Rachel Nordlinger, Dennis Norris, Jane Simpson, and Pelle Soderstrom. We are grateful to Handling Editor Dana Basnight-Brown, to three anonymous reviewers, and to Patricia Clancy, for lucid critical comments.

Author contributions H. S. Sarvasy conceived of the project, helped design the experiment, conducted the experiment, processed and coded the data, and participated in writing and revising the paper. A. M. Morgan created statistical models, conducted analyses, and participated in writing and revising the paper. J. Yu designed and programmed the experiment, conducted initial analyses, and drafted the experimental methods section. V. S. Ferreira consulted on framing and analyses, and contributed to revisions. S. Momma consulted on framing and analyses, and contributed to revisions.

Funding Funding was received from the Australian Research Council (Grants CE140100041 and DE180101609), and from the MARCS Institute for Brain, Behaviour and Development at Western Sydney University.

Declarations

Ethics Ethics approval was granted by the Australian National University, approval number 2013/055.

Informed consent Informed consent was obtained from all individual participants included in the study.

The authors affirm that human research participants provided informed consent for publication of the image in Fig. 1.

Data availability/open practices statement The data sets generated during and/or analyzed during the current study are available in the “Nungon switch reference: Eyetracking during comprehension and production” repository (<https://osf.io/wp47e/>). Neither experiment was preregistered.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Altmann, G. T. (2004). Language-mediated eye movements in the absence of a visual world: The ‘blank screen paradigm’. *Cognition*, 93(2), B79–B87.
- Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, 73(3), 247–264.
- Altmann, G. T., & Kamide, Y. (2004). Now you see it, now you don’t: Mediating the mapping between language and the visual world. In *The interface of language, vision, and action: Eye movements and the visual world*, 347–386.
- Altmann, G. T., & Kamide, Y. (2009). Discourse-mediation of the mapping between language and the visual world: Eye movements and mental representation. *Cognition*, 111(1), 55–71.
- Altmann, G. T., & Mirković, J. (2009). Incrementality and prediction in human sentence processing. *Cognitive Science*, 33(4), 583–609.
- Baayen, R. H., Davidson, D., & Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390–412.
- 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(3), 255–278.
- Beattie, G. (1980). Encoding units in spontaneous speech: Some implications for the dynamics of conversation. In H.W. Dechert, M. Raupach (Eds.), *Temporal variables in speech* (pp. 313–144). Mouton.
- Bock, K., & Ferreira, V. (2014). Syntactically speaking. In M. Goldrick, V. Ferreira, & M. Miozzo (Eds.), *The Oxford Handbook of Language Production* (pp. 21–46). Oxford University Press.
- Boersma, P., & Weenink, D. (2019). Praat: Doing phonetics by computer [Computer a] Version 6.1.08. <http://www.fon.hum.uva.nl/praat/>
- Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review*, 113(2), 234.
- Clifton, C., & Frazier, L. (1989). Comprehending sentences with long-distance dependencies. In G. N. Carlson & M. K. Tanenhaus (pp. 273–317). Springer.
- DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, 8(8), 1117–1121.
- Dobrin, L. (2008). From linguistic elicitation to eliciting the linguist. *Language*, 84(2), 300–324.
- Dooley, R. A. (2010). *Exploring clause chaining*. SIL Electronic Working Papers.
- Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences*, 32(5), 429–48.
- Ferreira, F. (1991). Effects of length and syntactic complexity on initiation times for prepared utterances. *Journal of Memory and Language*, 30, 210–233.
- 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(1), 57–84.
- Ferreira, F., & Swets, B. (2005). The production and comprehension of resumptive pronouns in relative clause “island” contexts. In A. Cutler (Ed.), *Twenty-first century psycholinguistics: Four cornerstones* (pp. 263–278). Routledge.
- Ford, M. (1982). Sentence planning units: Implications for the speaker’s representation of meaningful relations underlying sentences. In J. Bresnan (Ed.), *The mental representation of grammatical relations* (pp. 797–827). MIT Press.
- Ford, M., & Holmes, V. M. (1978). Planning units in sentence production. *Cognition*, 6, 35–53.
- Garrett, M. F. (1975). The analysis of sentence production. In G. Bower (Ed.), *Psychology of learning and motivation* Vol 9, pp. 505–529. Academic Press.
- Garrett, M. F. (1982). Production of speech: observations from normal and pathological language use. In A. W. Ellis (Ed.), *Normality and pathology in cognitive functions* (pp. 19–76). Academic Press.

- Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science*, 11(4), 274–279.
- Haiman, J. (1980). *Hua: A Papuan Language of the Eastern Highlands of New Guinea*. Amsterdam: John Benjamins.
- Haiman, J., & Munro, P. (1983). *Switch Reference and Universal Grammar*. Amsterdam: John Benjamins.
- Hanne, S., Burchert, F., De Bleser, R., & Vasishth, S. (2015). Sentence comprehension and morphological cues in aphasia: What eye-tracking reveals about integration and prediction. *Journal of Neurolinguistics*, 34, 83–111.
- Huettig, F. (2015). Literacy influences cognitive abilities far beyond the mastery of written language. In *Adult literacy, second language, and cognition*. Centre for Language Studies.
- Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, 137(2), 151–171.
- Kamide, Y., Altmann, G. T., & Haywood, S. L. (2003a). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, 49(1), 133–156.
- Kamide, Y., Scheepers, C., & Altmann, G. T. (2003b). Integration of syntactic and semantic information in predictive processing: Cross-linguistic evidence from German and English. *Journal of Psycholinguistic Research*, 32(1), 37–55.
- Longacre, R. E. (1985). Sentences as combinations of clauses. In T. Shopen (Ed.), *Language typology and syntactic description, Volume II: Complex constructions* (pp. 235–286). Cambridge University Press.
- Lukyanenko, C., & Fisher, C. (2016). Where are the cookies? Two- and three-year-olds use number-marked verbs to anticipate upcoming nouns. *Cognition*, 146, 349–370.
- McDonald, J. L., Bock, K., & Kelly, M. H. (1993). Word and world order: Semantic, phonological, and metrical determinants of serial position. *Cognitive Psychology*, 25(2), 188–230.
- Meir, N., Parshina, O., & Sekerina, I. A. (2020). The interaction of morphological cues in bilingual sentence processing: An eye-tracking study. In M. M. Brown & A. Kohut (Eds.), *Proceedings of the 44th Boston University Conference on Language Development* (pp. 367–389). Cascadilla Press.
- Meyer, A. S. (1996). Lexical access in phrase and sentence production: Results from picture-word interference experiments. *Journal of Memory and Language*, 35, 477–496.
- Mitsugi, S. (2017). Incremental comprehension of Japanese passives: Evidence from the visual-world paradigm. *Applied Psycholinguistics*, 38(4), 953–983.
- Momma, S., & Ferreira, V. S. (2019). Beyond linear order: The role of argument structure in speaking. *Cognitive Psychology*, 114, 101228.
- Momma, S., Slevc, L. R., & Phillips, C. (2016). The timing of verb selection in Japanese sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42(5), 813.
- Momma, S., Slevc, L. R., & Phillips, C. (2018). Unaccusativity in sentence production. *Linguistic Inquiry*, 49(1), 181–194.
- Momma, S., & Yoshida, M. (2021). *Planning multiple dependencies in sentence production*. Presentation at the 34th annual CUNY conference on human sentence processing.
- Mulak, K., Sarvasy, H., Tuninetti, A., Escudero, P. (2021). Word learning in the field: a case study from Papua New Guinea. *PLOS ONE* <https://doi.org/10.1371/journal.pone.0257393>
- Pawley, A., & Snyder, F. H. (2000). The one-clause-at-a-time hypothesis. In *Perspectives on fluency*. University of Michigan Press.
- Pizarro-Guevara, J. S., & Wagers, M. (2020). The predictive value of Tagalog voice morphology in filler-gap dependency formation. *Frontiers in Psychology*, 11, 517.
- Lombardi, L., & Potter, M. C. (1992). The regeneration of syntax in short term memory. *Journal of Memory and Language*, 31(6), 713–733. [https://doi.org/10.1016/0749-596X\(92\)90036-W](https://doi.org/10.1016/0749-596X(92)90036-W)
- Sarvasy, H. S. (2017). *A Grammar of Nungon: A Papuan language of northeast New Guinea*. : Brill.
- Sarvasy, H. S. (2019). The root nominal stage: a case study of early Nungon verbs. *Journal of Child Language* 46(6), 1073–1101.
- Sarvasy, H. S. (2020). Acquisition of clause chaining in Nungon. *Frontiers in Psychology*, 11, 1456. <https://doi.org/10.3389/fpsyg.2020.01456>
- Sarvasy, H. S. (2021). Quantifying clause chains in Nungon texts. *Studies in Language*. <https://doi.org/10.1075/sl.19058.sar>
- Sarvasy, H. S. & Choi, S. (2020). Beyond the two-clause sentence: Acquisition of clause chaining in six languages. *Frontiers in Psychology*.
- Smith, M. & Wheeldon, L. (1999). High level processing scope in spoken sentence production. *Cognition*, 73, 205–246.
- van Gijn, R., & Hammond, J. (2016). *Switch Reference 2.0*. Amsterdam: John Benjamins.
- Wagers, M. W., Lau, E. F., & Phillips, C. (2009). Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language*, 61(2), 206–237.
- Wagner, V., Jescheniak, J. D., & Schriefers, H. (2010). On the flexibility of grammatical advance planning during sentence production: Effects of cognitive load on multiple lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(2), 423–440. <https://doi.org/10.1037/a0018619>
- Wijnen, F. (1990). The development of sentence planning. *Journal of Child Language*, 17, 651–675.
- Wray, A., & Grace, G. W. (2007). The consequences of talking to strangers: Evolutionary corollaries of socio-cultural influences on linguistic form. *Lingua*, 117(3), 547–578.
- Yoshida, M. (2004). *Relative clause prediction in Japanese*. Unpublished manuscript.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.