Language and Cognition (2023), 1–26
doi:10.1017/langcog.2023.49

ARTICLE

Past and future time reference processing teased apart in Paiwan, an endangered Formosan language

Aymeric Collart and Elizabeth Zeitoun

Institute of Linguistics, Academia Sinica, Taipei, Taiwan
Corresponding author: Aymeric Collart; Email: aymeric.collart@gmail.com

(Received 19 October 2022; Revised 18 September 2023; Accepted 19 September 2023)

Abstract

How the time reference of a sentence is processed based on the grammatical marking of the verb has already been explored in several languages with grammatical tense and aspect. It can also be grammatically expressed according to the reality status of the event (whether the event exists in time, *realis* mood, or not, *irrealis*). This study reports results from an acceptability judgment experiment in Paiwan, a Formosan language which exhibits a realis-perfective/irrealis distinction. By placing realis-perfective and irrealis markers after deictic past or future time adverbs and manipulating the grammaticality of the sentences, we asked which temporal concord violation (i.e., realis-perfective or irrealis) was harder to detect. The temporal concord violation of the realis-perfective marker induced greater processing difficulties (interactions between time reference and mood marking revealed lower accuracy rates and longer reaction time), but not the irrealis marker, in line with previous hypotheses. These processing difficulties may be partly due to the Paiwan realis mood marker which also encodes perfective aspect meaning. The reanalysis of the design of previous studies indicates that the interaction with perfective aspect also led to additional processing cost, suggesting that perfective aspect marking plays a crucial role in the processing of time reference.

Keywords: time reference; realis/irrealis; perfective aspect; Paiwan; acceptability judgment; PADILIH

1. Introduction[1]

It is well known that language processing research heavily relies on a small number of languages which mainly belong to the Indo-European language family (Anand et al.,

[1]The glosses used in this paper include: AV, actor voice; COS, change of state; DIR, directional; GEN, genitive; IRR, irrealis; LIG, ligature; NEG, negation; NOM, nominative; OBL, oblique; PAT.NMLZ, patient nominalization; PFV, perfective; PL, plural; PN, proper nouns; RED, reduplication; SG, singular; STAT, stative verb; UV,

2011; Kidd & Garcia, 2022). This means that generalizations of processing patterns mostly draw on the linguistic characteristics shared by Indo-European languages (Haspelmath, 2001), implying that features found in typologically different languages are less likely to be taken into account in these processing models. Heinrich et al. (2010, p. 29) and Majid and Levinson (2010, p. 103) use the acronym 'WEIRD' (languages spoken in 'Western, Educated, Industrialized, Rich and Democratic' societies) to describe this bias, claiming that "WEIRD languages have misled us".[2] As an example of how a non-WEIRD language can contribute to our knowledge of language processing, Maia (2021) shows how empathetic deixis marking in Karajá (belonging to the Macro-Jê language family) influences pronoun resolution processing, and this interaction was found by conducting experiments on a non-WEIRD language which exhibits morphosyntactic characteristics absent in Indo-European languages.

There are other morphosyntactic constructions displaying important crosslinguistic differences, such as the expression of the time reference of a sentence which can be morphosyntactically encoded with tense, aspect, mood, or modality (Bhat, 1999; Dahl, 1985). Models based on behavioral and neuroimaging studies have also been proposed regarding how time reference is processed, such as the *Past Discourse Linking Hypotheses* (PADILIH), assumed to be valid crosslinguistically (Bastiaanse et al., 2011; Dragoy et al., 2012). However, as we will show in the subsequent sections, the PADILIH also relies on a few languages which only cover some of the possible ways to express the time reference of a sentence. It thus raises doubt about the crosslinguistic validity of the current version of the PADILIH.

This paper aims at testing the claims made by the PADILIH by investigating the processing of time reference at the sentence level in Paiwan (Austronesian language of Taiwan, conventionally referred to as a Formosan language), which exhibits morphosyntactic characteristics not found in the languages used to support the PADILIH. Section 2 sets the background of this study, including the theoretical assumptions and the definitions of tense, aspect, and mood (Section 2.1), a review of previous behavioral studies investigating the processing of the temporal reference of a sentence with deictic time adverbs and grammatical marking showing how these studies relate to the PADILIH (Section 2.2), and a short introduction to Paiwan grammar, in particular its grammatical means to express time reference (Section 2.3). Section 3 lays the rationale of the present study, including the experiment design and the hypotheses and predictions. The methodology is described in Section 4, and the results are reported in Section 5. In Section 6, we discuss the results according to the main research question (i.e., is the processing pattern of time reference in Paiwan predicted by the PADILIH?) along with the insights that the results of our Paiwan experiment brings to the processing of time reference in general.

2. Time, language, and the processing of time reference

2.1. Linguistic background on the expression of time reference

The time reference of a sentence, or the perception of when the event mentioned in the sentence occurs, can be expressed with different means. Lexical time adverbs,

undergoer voice; UVC, undergoer voice—circumstantial; UVL, undergoer voice—locative; UVP, undergoer voice—patient.

[2]Other acronyms have been proposed in the literature, such as 'LOL' (Dahl, 2015), standing for languages which are "literate, official, and with lots of users".

https://doi.org/10.1017/langcog.2023.49 Published online by Cambridge University Press

such as deictic time adverbs like *yesterday* or *tomorrow* can be used to set up a specific temporal frame. On the grammatical side, time reference can be encoded with different categories, and in particular tense, aspect, and mood. Tense is defined as the grammaticalization (that is the morphological marking) of the localization of an event in time; it divides between absolute tense (time of speech as temporal anchor) and relative tense (time of another event taken as reference and not necessarily the time of speech; Comrie, 1985, pp. 14–17).

Grammatical aspect concerns the constituency of the event itself (Comrie, 1976, p. 3). An event can be seen in its entirety, including both the initial and final temporal point of the event; it corresponds to 'perfective' aspect. Contrastively, imperfective aspect focuses on the internal constituency of the event (i.e., a time interval found between the initial temporal point and the final temporal point of the event), with two subtypes of imperfective aspect: (1) habitual aspect, and (2) progressive aspect. In many languages without tense marking, sentences involving perfective aspect are often interpreted as exhibiting past time reference, and with imperfective aspect non-past time reference (Li, 2012; Smith & Erbaugh, 2005).

Another category related to the expression of temporal reference is mood, in particular the distinction between realis and irrealis (Bhat, 1999; Palmer, 2001). While realis is used to present an event taken as being actualized or occurring at the moment, irrealis is concerned with unrealized events (de Haan, 2012; Elliott, 2000; see also von Prince et al. (2022) for a time-branching treatment of the crosslinguistic differences of realis and irrealis). Irrealis covers several interpretations, prototypically non-factual. Correspondingly, irrealis can occur in constructions such as counterfactual constructions, the apodosis of conditionals, imperatives, and commands, as well as future/projective constructions, among others (Elliott, 2000; Plungian, 2005). In terms of temporal reference, realis is often associated with past time reference since past events are prototypically perceived as realized and fixed, and irrealis with future time events, since they are hypothesized and imagined and therefore non-factual. This relationship is not absolute, as realis can be used to express future time reference as long as it is very likely to occur, i.e., this future event is not hypothesized (de Haan, 2012, p. 23; Elliott, 2000, p. 68), and irrealis can be found in past contexts which do not denote an actual or fixed situation as in counterfactual ones.

The use of deictic time adverbs may interact with grammatical morphemes. For example, in English, a deictic past time adverb generally requires a past tense marking on the verb. This phenomenon is referred to as 'tense agreement'. Some other studies (e.g., Biondo et al., 2021) employ the term 'temporal concord' instead of 'tense agreement' since a temporal mismatch can also occur between a deictic time adverb and a verbal morpheme which is not a tense marker. This is the case for example of Mandarin, where the use of the grammatical aspect markers *-le* and *-guo* are ungrammatical after a deictic future time adverb in numerous sentences (Collart & Chan, 2021).

2.2. Processing time reference: PADILIH and insights from behavioral studies

The processing patterns of time reference have been investigated in different languages, such as English (Bastiaanse et al., 2011; Faroqi-Shah & Dickey, 2009), Dutch (Baggio, 2008; Bos et al., 2013; Dragoy et al., 2012), Italian (Biondo et al., 2021), Mandarin Chinese (Bastiaanse et al., 2011; Collart & Chan, 2021), and Spanish (Biondo et al., 2022), using a variety of types of investigation (e.g., patient studies,

https://doi.org/10.1017/langcog.2023.49 Published online by Cambridge University Press

behavioral studies, ERP or fMRI). These studies point out an asymmetry when processing past and non-past temporal frames. Evidence from the behavioral studies conducted with healthy population available in the literature is summarized in turn below, corresponding to Biondo et al. (2022), Collart and Chan (2021), Dragoy et al. (2012), and Faroqi-Shah and Dickey (2009).[3]

Faroqi-Shah and Dickey (2009) investigated the processing of tense agreement mismatch in English with sentences involving different time frames which could be morphologically encoded with verbal suffixed or auxiliary verbs. See (1–3).

(1) Last year / *Next year, my sister lived in New Hampshire. (past tense, suffix)
(2) These days / *Last year, my sister lives in Boston. (present tense, suffix)
(3) Next year / *Last year, my sister will live in Boston. (future-tense-like, auxiliary verb)

In addition to measuring the accuracy of the grammaticality judgments of these sentences as well as the reaction time with aphasic people, Faroqi-Shah and Dickey (2009) also included a control group of healthy participants on which we focus in the following. Faroqi-Shah and Dickey (2009, p. 106) report that the reaction time of the grammaticality judgments of sentences with past tense and future tense verbs was higher than with present tense verbs, indicating difficulties associated with these two tense forms when "translating conceptual-semantic information into tense morphology". Nevertheless, we note that the accuracy of the judgments of present-tense sentences was lower than of other tense forms, maybe reflecting a trade-off between reaction time and judgment task, but Faroqi-Shah and Dickey (2009) do not discuss this point. In addition, concluding that judging past-tense and future-tense verbal forms are both more difficult to process than present tense as Faroqi-Shah and Dickey (2009) do might be too hasty since there may have been other confounding factors, such as (1) the different morphosyntactic encoding of past tense (verbal suffix) and future tense (auxiliary verb) in English, (2) the analysis pipeline which does not distinguish between grammatical and ungrammatical sentences, and (3) the fact that the auxiliary verb *will* may not be completely considered as a future tense marker.

Dragoy et al. (2012) conducted a study of the same kind as Faroqi-Shah and Dickey (2009) in Dutch. While the main focus of this study is the processing of time reference when the verb is inflected with tense markers using the ERP technique, Dragoy et al. (2012) also ran a behavioral experiment in which the participants had to react as soon as they detected an anomaly in the sentence they were reading. A temporal concord violation was created by placing a verb inflected for past tense after a present-time adverb, and also by placing a present tense verb after a past-time adverb (sentences in which the tense morpheme agrees with the time adverb were also used in this experiment). Crucially, unlike Faroqi-Shah and Dickey (2009), the same kind of morphosyntactic encoding was used across the experimental conditions. The temporal concord violation of the past-tense verbs and present-tense verbs

[3]These studies assume that the judgment ratings reflect the comprehension process of the temporal concord of the sentences in question, while these ratings could also be taken as just reflecting a non-error/error detection process. In other words, the results inform about the complexity of the detection of the temporal concord violation rather than the processes by which the temporal concords are computed.

https://doi.org/10.1017/langcog.2023.49 Published online by Cambridge University Press

showed dissimilarities in both terms of accuracy and reaction time. The accuracy rates of the ungrammatical sentences were 90% for sentences with a present tense verb, but only 74% with a past tense verb. The participants also detected the violation of present tense faster than past tense (1370 ms vs. 1760 ms). Dragoy et al. (2012) take these differences as reflecting qualitatively different processing mechanisms. These two studies provide evidence concerning the processing disadvantage of past-time reference compared to present-time reference. The experiment conducted by Biondo et al. (2022) in Spanish further suggests that processing past time reference yields greater difficulties than future time reference as well. In this study, Biondo et al. (2022) asked the participants to categorize verbs as inflected for past or future tense as fast as possible. The participants required more time to categorize past-tense verbs than future-tense verbs, and Biondo et al. (2022) interpret this difference as an indication of enhanced processing difficulties for past-time reference.

Other studies using the same experimental paradigm further show that not all the markers expressing past time reference are equally hard to process. Collart and Chan (2021) explored the neuronal processing of two aspect morphemes in Mandarin Chinese, *-le* and *-guo*. Linguists generally agree that the aspect morpheme *-le* expresses perfective aspect meaning, while *-guo* denotes experiential or phasal aspect (Li, 2012; Li & Thompson, 1981). The behavioral results reported in Collart and Chan's (2021) study indicate that judging the perfective marker *-le* in the temporal concord violation context-induced lower accuracy rates than its grammatical counterpart. Importantly, this difference was not found for *-guo*. These results suggest that in addition to the difference of processing difficulties between past and present-time reference when expressed with tense markers reflected in lower accuracy rates and/or longer reaction time, further mechanisms may influence this difficulty of processing, such as the perfective aspect.

The asymmetry between the processing of past- and present-time reference expressed with past and present tense has been formalized within the PADILIH (Bastiaanse, 2018; Bos et al., 2013; Dragoy et al., 2012). The PADILIH draws on two linguistic theories: the time-relational theory of tense of Reichenbach (2005/1947), which corresponds to the definition of tense by Comrie (1985), and the theories of Partee (1973) and Zagona (2003) on the other hand. In the latter theories, tense features are taken as being similar with pronominal features: the time at which an event occurs must find its time of reference, just like a pronoun must find the antecedent NP to which it refers. When a verb is marked for present tense, the time of the event coincides with the time of speech. Contrastively, the time of the event does not coincide with the time of speech for the case of past tense, such that an additional linking between the time of the event and the time of speech is required. It is argued under the PADILIH that this linking is discourse-based. Therefore, processing past tense would require more cognitive load than present tense due to this additional linking process. The extra effort would be reflected in the difference of reaction time and/or accuracy rates between past tense and present tense in Dragoy et al. (2012) and Faroqi-Shah and Dickey (2009).

More remarks concerning the assumptions made by the PADILIH are necessary. First, the PADILIH, in its original version, assumes a dichotomy between past-time reference and non-past-time reference, hence merging present- and future-time reference together. Under the PADILIH, future-time reference is assumed to be linked with present-time reference. Indeed, since future events have not happened yet, they are considered as a projection of the representation of the event as thought at

https://doi.org/10.1017/langcog.2023.49 Published online by Cambridge University Press

the present time (Bastiaanse et al., 2011, p. 657). Furthermore, since a future event is projected from the present moment, this entails that there is no discursive operation when establishing the temporal relation between the time of speech and the time of the future event. In other words, future events are also considered as being locally rather than discursively processed. Second, the PADILIH assumes that such a linking is not restricted to *tense* as a grammatical category *per se*, but to *time reference* as a concept, such that the PADILIH also holds for all the formal expressions of time reference, such as aspect and mood. Consequently, Bastiaanse et al. (2011) also claim that the PADILIH is cross-linguistically valid. However, the studies reviewed in this section are mainly concerned with the processing of time reference when expressed with tense marking. In other words, they are informative for only part of the picture, hence the importance to conduct similar studies with languages using other means to express time reference, as pointed out by Bastiaanse (2018). Paiwan appears to be a suitable candidate for trying to replicate the results found in Dragoy et al. (2012) or Faroqi-Shah and Dickey (2009), and test the PADILIH because of the way time reference is expressed in Paiwan grammar which differs from these languages, as exposed in the following section.

2.3. Presentation of Paiwan: sketch grammar and the expression of time reference in Paiwan

With a population of 106,016 as of August 2023 (Council of Indigenous Peoples, 2023), Paiwan is the second largest ethnic group of Taiwan. The Paiwan reside in the south of the island and form 125 communities in villages situated across Pingtung and Taitung Counties.

Paiwan is an agglutinative language, with a rich morphology. It exhibits two productive morphological processes, affixation and reduplication, which may occur independently or combine together (see Chang, 2006; Huang, 2012; Ng, 2023).[4] It is a predicate-initial language, with the subject occurring just after the predicate or in sentence-final position. The most important grammatical features concern voice marking on the verb and case marking on noun phrases. We introduce them in turn below.

Paiwan makes a basic distinction between dynamic verbs, usually marked by <*em*> (and its allomorphs) in actor voice, for example, *tekel* 'drink' (4a), and stative verbs, marked by Ø, for example, *tjengelay* 'like' and to a lesser extent *ma-*, for example, *ma-lum* 'be ripe' (4b).[5]

(4) Paiwan

a. **na=tekel** a kama tua zaljum.
PFV=<AV>drink NOM father OBL water
'Father drank water.'

[4]Paiwan is one of the best documented Formosan languages, with studies covering all aspects of grammar, including phonetics, phonology, morphology, and syntax. The documentation includes dictionaries (e.g., Ferrell, 1982), grammars (e.g., Chang, 2006; Ng, 2023) and texts (e.g., Egli, 1989; Ferrell & Tjakisuvung, 2023).

[5]There are 23 consonants and 4 vowels in Paiwan, including: *p* /p/, *b* /b/, *t* /t/, *d* /d/, *dr* /ɖ/, *tj* /c/, *dj* /ɟ/, *k* /k/, *g* /g/, *q* /q/, *v* /v/, *s* /s/, *z* /z/, *c* /ʦ/, *m* /m/, *n* /n/, *ng* /ŋ/, *l* /l̪/, *lj* /ʎ/, *r* /r/, *w* /w/, *y* /j/, (*'* /ʔ/) (the glottal stop occurs in very few words), *i* /i/, *u* /u/, *e* /ə/, and *a* /a/.

https://doi.org/10.1017/langcog.2023.49 Published online by Cambridge University Press

b. **na=ma-lum**=anga azua vangalj nua djaraljap.
PFV=STAT-ripe=COS NOM.that fruit GEN banyan
'The figs were already ripe.' (Ferrell & Tjakisuvung, 2023), B7, Ln 14)

Both types of verbs can take voice affixes. The voice system is based on a dichotomy between AV 'actor voice' and UV 'undergoer voice' which further subsumes UVP 'undergoer voice—patient', UVL 'undergoer voice—locative' and UVC 'undergoer voice—circumstantial'. This voice system interacts closely with mood (indicative and non-indicative) and aspect (perfective and imperfective). The indicative mood subsumes declarative and interrogative clauses and non-indicative mood, imperative, hortative and optative (Ross, 1995). The indicative mood further divides into realis and irrealis. When verbs only carry voice affixes (AV **, UVP *-en*, UVL *-an*, and UVC *si-*) with no other co-occurring temporal adverbs or aspectual markers, there is a clear understanding that the event has already happened or is happening, depending on the discourse context (Zeitoun et al., 1996). That is, the realis is unmarked. Examples (5a–b) illustrate verbs marked as realis AV and UVP in the indicative mood.

(5) Paiwan
a. **kun** tua k<in>im-an a pa-tje-cungal.[6]
<AV>skirt OBL <PFV>find-LOC.NMLZ LIG DIR-down.to-knee
'She put on a skirt (the sparrows) had found (that covered her body) from the feet to the knees.' (Ferrell & Tjakisuvung, 2023), B2, Ln 124)
b. **sevalj-en** nuazua ti kalji *aya* vaik *aya.*
carry-UVP GEN.that PN Kalji leave
'Kalji picked her up and they started out.' (Ferrell & Tjakisuvung, 2023), B2, Ln 20)

The progressive and the continuous/repetitive aspects are lumped together, i.e., they are marked by CVCV-reduplication and subsumed as realis imperfective. In AV clauses, the perfective is marked by *na=* (6a) and in UV clauses by the infixation of *<in>* (6b). In both instances, the event is treated as having already happened.

(6) Paiwan
a. **na=kun** tua k<in>iman a pa-tje-cungal
PFV=<AV>wear. OBL <PFV:PAT.NMLZ>extend LIG DIR-down.to-
skirt knee
'She wore a skirt that reached to her knees.' (Ferrell & Tjakisuvung, 2023), B1, Ln 194)
b. nia=**p<in>icasaw**=anga tja=sicauvan lje-teku=anga.
we.GEN=<PFV:PAT.NMLZ>put=COS our=blanket go-below=COS
'We already put (the blanket) below.' (Ferrell & Tjakisuvung, 2023), B2, Ln 193)

In the irrealis, the verb is cliticized by *uri=*. A verb marked as irrealis may refer to a projective event, as in (7a–b), but can also encode a hypothetical or conditional event (the notions of past or future can only be retrieved from the contexts; Chang, 2006).

[6]There is no pronominal clitic for the third person (singular and plural) in Paiwan.

https://doi.org/10.1017/langcog.2023.49 Published online by Cambridge University Press

(7) Paiwan
a. “ai=anga! **uri=qudjalj**” aya
ugh=COS IRR=<AV>rain say
“Ugh, it is going to rain.” (Ferrell & Tjakisuvung, 2023), B2, Ln 124)
b. “ai=anga! lja-qali-an **uri=qudjalj-an**=itjen” aya
ugh=COS PL-friend IRR=rain-UVL=we.NOM say
“Oh-oh, friends, we’re about to get rained on!” he said. (Ferrell & Tjakisuvung, 2023), B1, Ln 107)

This means that even if the default temporal interpretation of the irrealis clitic *uri=* is future, it can also be found in past contexts as long as they are hypothetical. See (8) (Chang, 2006, p. 183).

(8) Paiwan
nu **uri=sane-kava=[a]men** tua nia=kava,
if IRR=<AV>make-dress=we.NOM OBL our=dress
ini=ka pu-tjuva~tjuvak tua kava nua va~vaya~vayan.
NEG1=NEG2 have[AV]-RED~sea.shell OBL dress GEN RED~female
‘If (we) would like to make our clothes, we did not put sea shells on women’s clothes.’ (Chang, 2006, p. 183)

Paiwan’s verbal morphology is summarized in Table 1.

Paiwan case markers are divided into two sets: (1) personal proper nouns, with a further distinction between singular and plural, and (2) common nouns. There are three cases: nominative, genitive, and oblique. In AV clauses, an actor is marked with the nominative case, and the patient as oblique, as shown in Table 2. In UV clauses, the non-subject actor is marked with genitive, and the undergoer, which is the subject, is marked by the nominative. Note that any argument (patient/theme, instrument, recipient, location) may be marked as nominative if selected as the subject of the clause.

Table 1. Paiwan verbal morphology (based on Huang (2012) and Ng (2023))

				UV		
			AV	UVP	UVL	UVC
Indicative	Realis	Neutral	**	*-en*	*-an*	*si-*
		Imperfective	** + RED	*-en* + RED	*-an* + RED	*si-* + RED
		Perfective	*na=*	*<in>*	*<in>-an*	*s<in>i-*
	Irrealis		*uri=*	*uri=* + *-en*	*uri=* + *-an*	*uri=* + *si-*
Non-indicative		Imperative	*-u*	*-u*	*-i*	*-an*
		Hortative	*-i*	–	–	–
		Optative	–	*-aw*	*-ay*	*si-…-an*
		Dependent	Ø	*-i*	*-i*	*-an*

Table 2. Paiwan case markers

		Nominative	Genitive	Oblique
Personal proper noun	singular	*ti*	*ni*	*tjai*
	plural	*tia*	*nia*	*tjaya*
Common noun		*a*	*n(u)a*	*t(u)a*

https://doi.org/10.1017/langcog.2023.49 Published online by Cambridge University Press

The examples given in (9) and (10) illustrate the distribution of case markers for personal proper nouns (in the singular) and common nouns in the nominative, oblique, and genitive.

(9) Paiwan
a. uri=ljepljep **ti** tjuku tua nemanga.
IRR=<AV>tidy NOM Tjuku OBL thing
'Tjuku will tidy things.'
b. na=kuyac=aken **tjai** kuljelje.
PFV=<AV>curse=I.NOM OBL Kuljelje
'I cursed Kuljelje.'
c. ki-anema a k<in>sa **ni** kama?
after.all-what NOM <PFV:PAT.NMLZ>cook GEN father
'What will father cook after all?'

(10) Paiwan
a. na=v<en>eric **a** kina **tua** cemel.
PFV=<AV>throw NOM mother OBL garbage
'Mother threw the garbage.'
b. uri=laing-en azua va~vaya~vayan **nua** uqaljay.
IRR=chase-UVP NOM.that RED~female GEN man
'That woman will be chased by the man.'

There are two temporal markers in Paiwan found on temporal adjuncts: *ka-*, which refers to a past event (realis), as in (11a), and *nu-*, which indicates a future event (irrealis), as in (11b) (Sung, 2006; Zeitoun et al., 1996). These same markers are used in 'past/future-when'-clauses.

(11) Paiwan
a. **katiaw** na=rata a aljak tua mudinga.
yesterday PFV=<AV>wash NOM child OBL face
'Yesterday, the child washed his face.'
b. **nutiaw** uri=rawraw a kina tua kava.
tomorrow IRR=<AV>wash NOM mother OBL clothes
'Tomorrow, mother will wash clothes.'

3. The present study: experimental design and hypotheses

As mentioned above, the present study proposes to test the PADILIH with Paiwan (Formosan language) by investigating the processing of the temporal concord between deictic time adverbs and the grammatical marking of realis mood/perfective aspect clitic *na=* and the irrealis clitic *uri=*. Indeed, previous studies (Biondo et al., 2022; Dragoy et al., 2012; Faroqi-Shah & Dickey, 2009) mainly focused on tense, but extending the results to time reference as these studies propose require insights from languages using other grammatical means to express time reference (Bastiaanse, 2018). In addition, the Mandarin Chinese experiment conducted by Collart and Chan (2021) suggests that judging the temporal concord violation of perfective aspect when used to express past time reference also induces processing difficulties, but this has never been tested in comparison with future time reference.

https://doi.org/10.1017/langcog.2023.49 Published online by Cambridge University Press

Based on these linguistic observations, we selected Paiwan as the language of research in the present study because of its grammatical properties regarding the expression of time reference. Paiwan is indeed suitable to set up an experiment using the same experimental paradigm as in the above-mentioned studies. The marker *na*= (perfective aspect constrained in realis mood context) is grammatical after a deictic past time adverb, but not after a deictic future time adverb. On the contrary, *uri*= (irrealis mood) is grammatical after a deictic future time adverb, but not after a deictic past time adverb.[7] The related experimental design inspired from these characteristics is given in Table 3.

Selecting Paiwan as the language of investigation, and the clitics *na*= and *uri*= in particular, allows us to state different hypotheses and predictions in order to test the PADILIH. Indeed, previous studies mainly focused on tense, which is mainly the time-relational expression of time reference, while mood is more related to the possible world semantics of the expression of time reference. This semantic difference generates different hypotheses:

1. The PADILIH claims that detecting the temporal concord violation of a past-time-related morpheme is harder than a non-past-time-related morpheme. Therefore, the PADILIH predicts that the ungrammaticality of *na*= after a deictic future time adverb (ungrammatical sentence) is harder to detect than its grammatical counterpart. By contrast, it also predicts that detecting the ungrammaticality of *uri*= after a past time adverb (ungrammatical sentence) is not harder than its grammatical counterpart (i.e., *uri*= after a future time adverb). These can be reflected by a lower accuracy rate and/or longer reaction time for *na*=, but nor for *uri*= .
2. On the other hand, it can be argued that irrealis events are harder to process than realis events since we need to imagine and project events that have not yet happened (Lyons, 1977). Smith and Erbaugh (2005) proposed the "default deictic pattern" (a pragmatic explanation similar to the "principle of Maximize Presupposition" in Mucha (2015) accounting for the temporal interpretation of sentences without tense), assuming that future time reference is harder than past because of its uncertainty. Some theories of mental time travel (Suddendorf, 2010; Suddendorf et al., 2009) have also considered that future time reference processing involves greater difficulties than past because of this characteristic.

   Under this view, we hypothesize that detecting the temporal violation of *uri*= should be harder when compared with its congruent counterpart, as it supposes to imagine a hypothetical event. This should be reflected by a lower accuracy rate and/or a longer reaction time for *uri*= in the ungrammatical

[7]We checked this grammaticality pattern at different steps of the elaboration of this study. First, we went through a literature review to make sure that the grammaticality pattern was stable based on the elicited sentences and linguistic analyses. Then, we collaborated with a linguistic-trained consultant, who is a native speaker of Paiwan from the village where the experiment took place, in order to confirm the stability of the grammaticality pattern when adjusting other parameters (i.e., the category or the meaning of the verbs). Finally, we went to the village several months before the experiment to interview other native speakers. They confirmed the pattern, corroborating past linguistic analyses and the discussions with the consultant. Therefore, when we started our experiment, we were confident enough to conclude that this is a categorical grammaticality pattern.

https://doi.org/10.1017/langcog.2023.49 Published online by Cambridge University Press

Table 3. Experimental design of the current study

Time reference	Marker	Example sentence					
Past-time reference	na=	katiaw	na=kan	a	kakedrian	tua	vurasi.
		yesterday	NA=<AV>eat	NOM	child	OBL	sweet.potato
		'Yesterday, the child ate sweet potatoes.'					
	uri=	*katiaw	uri=kan	a	kakedrian	tua	vurasi.
		yesterday	URI=<AV>eat	NOM	child	OBL	sweet.potato
		*'Yesterday, the child will eat sweet potatoes.'					
Future-time reference	na=	*nutiaw	na=kan	a	kakedrian	tua	vurasi.
		tomorrow	NA=<AV>eat	NOM	child	OBL	sweet.potato
		*'Tomorrow, the child ate sweet potatoes.'					
	uri=	nutiaw	uri=kan	a	kakedrian	tua	vurasi.
		tomorrow	URI=<AV>eat	NOM	child	OBL	sweet.potato
		'Tomorrow, the child will eat sweet potatoes.'					

condition when compared with its grammatical counterpart. Importantly, this additional processing difficulty should not be found for *na*=, since it does not denote an imagined event but an actualized event.

4. Methodology[8]

4.1. Participants

We recruited 23 participants (the data collection process took place in 2020). The recruitment process went through several stages. We went to the village several months before the data collection to meet and discuss with prospect participants. Because of logistic constraints, we set a limit of two weeks of data collection. We contacted potential participants based on the previous field trip, and they freely decided to take part in the experiment. Other participants willing to do the experiments were also recruited as long as they met the criteria. No sample size limit was set as we could not estimate how many participants could actually be recruited. Among the 23 recruited participants, two left the experiment before it ended. We also removed the data of three additional participants from the analyses because they did not focus on the experimental task (they expressly told the experimenters that they judged whether the words composing the sentences belong to the dialect of Paiwan spoken in their village). Therefore, the final number of participants included in the analyses is 18 (9 females and 9 males; $M = 60 \pm 12$ years old, SD = 10.1). They were all native speakers of Paiwan, and evaluated themselves as being fluent in Paiwan. Crucially, we paid attention to recruit the participants according to the dialect of Paiwan they were speaking in order to minimize dialectal differences in terms of phonology or word usage. Before the beginning of the experiment, all the participants signed a consent form approved by Academia Sinica. The participants were compensated with coupons after completing the experiment.

4.2. Material

We created 152 sentences with the help of a consultant, who had been trained to process and transcribe linguistic data for many years and who is used to collaborate with linguists. They consisted of the structure [Deictic time adverb – *na*=/*uri*= – Predicate – Nominative case – Noun – Oblique case – Noun]. Two types of deictic time adverbs were used: *katiaw* 'yesterday', giving a past time interpretation to the sentence, and *nutiaw* 'tomorrow', for a future time interpretation. Therefore, there were four types of sentences: (1) Past-Na, (2) Future-Na, (3) Past-Uri, (4) Future-Uri.

We chose the predicates according to several criteria: (1) dynamicity (no stative predicates were used), (2) syntactic transitivity (each predicate called for a subject and an object complement, ruling out intransitive and ditransitive predicates), and (3) relative concreteness. The predicates all appeared in the Actor Voice. We selected the subject nouns based on the following criteria: (1) animacy, (2) concreteness, and (3) natural link with the predicate. The object nouns were chosen for (1) concreteness and (2) naturalness of their occurrence with the lexical semantics of the preceding predicate and subject noun. The nouns consisted of common and proper nouns. We also made the choice to select the subject and object nouns according to their

[8]See Appendix A for general considerations on conducting experimental studies on endangered languages.

https://doi.org/10.1017/langcog.2023.49 Published online by Cambridge University Press

reference to the everyday life of the participants in the village, therefore excluding nouns referring to a more urban lifestyle which are less usual for the inhabitants of the village (e.g., 'computer'). By choosing this strict selection of the experimental material, the number of possible candidates was highly reduced, and we had to repeat some nouns across sentences. However, since the critical part of the sentence occurs at the beginning of the sentence (i.e., the predicate), the subject and object nouns have a relatively low impact on the experimental design, hence the choice to repeat the nouns to facilitate the comprehension of the sentences and not driving the attention of the participants on less common nouns. Nevertheless, we counterbalanced the occurrence of the repeated nouns between the conditions within each list to neutralize the repetition effect.

We created the experimental sentences according to the following procedure. First, we looked for sentences that met the criteria above described in the existing literature (Chang, 2006; Ferrell, 1982; Huang, 2012; Sung, 2006). We checked the sentences with several informants from the village in which the experiment took place. At the same time, we asked the informants to produce additional simple sentences (i.e., following the structure [Predicate (AV form) – Nominative case – Noun – Accusative case – Noun]) based on the predicates given to them. We then rechecked these sentences with other informants. The sentences including the time adverbs and the markers *na*=/*uri*= were then constructed and checked with a trained informant. Finally, all the constructed sentences were recorded in a sound-proof room at the Academia Sinica (Taipei, Taiwan) by the same informant. We did not recruit the informants involved in the material construction process for the experiment. The experimental sentences were distributed into four lists using a Latin square design, such that each experimental sentence appeared in the four conditions, but each participant only saw a part of them. Each participant only saw one list. The reason to adopt a Latin Square design is that it allowed us to counterbalance the sentences above several factors: (1) the frequency of use of the predicates and nouns, (2) the collocation of the subject noun with the object noun, as well as the subject noun with the predicate and the object noun with the predicate. There were 38 sentences for each condition in each list. The mean duration of the sentence by condition was: Past-Na: 4161 ms (SD = 355), Future-Na: 4146 ms (SD = 310), Past-Uri: 4130 ms (SD = 325), Future-Uri: 4143 ms (SD = 281). Overall, the difference of duration of the sentences between the conditions was not significant, implying that potential differences in the results were not due to the difference of duration ($F(3,604) = 0.243$, $p = .87$). We also checked whether the duration of the sentences differed between the experimental conditions once dispatched into four lists. The differences of duration were also not significant (main effect of Condition: $F(3,592) = 0.242$, $p = .87$; main effect of List: $F(3,592) = 0.435$, $p = .73$; Condition × List $F(9,592) = 0.788$, $p = .63$).

4.3. Procedure

Each participant completed the experiment individually in a quiet environment. They sat in front of a laptop and were told to relax. After signing the consent form, the experiment began with a practice session during which the experimenters explained the procedure of the experiment to the participants. The participants had the opportunity to go through the practice session more than one time, until

https://doi.org/10.1017/langcog.2023.49 Published online by Cambridge University Press

they were familiar with the experimental procedure. The experimental sentences were auditorily presented with speakers, and we adjusted the sound volume if necessary during the practice session. The participants had to judge the acceptability of the sentences as soon as they wished. To decrease the participants' cognitive burden, we used a binary task instead of a rating scale: the participants had to choose between "acceptable" and "unacceptable" responses. The participants had to press on an external numeric pad to provide their answers: the button '9' (on the right) corresponded to "acceptable", and '7' to "unacceptable". We also instructed them to hold the numeric pad with their two hands, such that they could press on '7' with their left thumb, and on '9' with their right thumb. We sticked colored dots on the buttons: a green sticker on '9', and a red sticker on '7'. Two colored icons were also displayed on the screen to remind the participants of which hand was corresponding to 'acceptable' and 'unacceptable'. We assumed that half of the experimental sentences would be acceptable, and the other half unacceptable, such that the number of "acceptable" and "unacceptable" answers was counterbalanced. The sentences were presented in a random order. The whole experiment was divided into four sections. Each section consisted of 38 sentences, and lasted for about four minutes. The whole experiment, including the explanation of the task, the practice trials, and the experimental session lasted for about 30 minutes. After completion, the experimenter debriefed the experiment with the participants and they completed a demographic questionnaire. The debriefing part consisted of individual discussions concerning, among others, the experiment itself. As noted by Wagers and Chung (2023), this was crucial as all the participants (except three of them) told us that they followed a 'rule' which mirrors the grammaticality pattern despite the fact that we paid extra attention not to mention this when providing the instructions. They also did not report any deviant sentences and were categorical regarding the judgment of the conditions. We also reminded the participants not to share the content of the experiments with other people in the village to prevent from an expectancy effect. Finally, as this is, to the best of our knowledge, the first psycholinguistic experiment in Paiwan, we were not sure whether this procedure might work with participants who were not used to it. Therefore, we conducted preliminary analyses of the results each time one of the participants completed the experiment in order to make sure that the procedure did not go wrong.

4.4. Analyses

We identified outlier trials based on the SD of the latency for each participant and each condition. Namely, we excluded trials for which the SD of the latency was 2.5 times below or above the mean latency from further analyses. This corresponded to 52 trials out of 2735, or 1.9% of the data. Then, the raw acceptability judgment responses were analyzed since this was the task that the participants had to perform. The aim of this first analysis was to check the tendency to accept or reject *na=* and *uri=* in certain temporal frames. We used logistic mixed-effects models to analyze the judgment acceptability responses using the 'lme4' package (Bates et al., 2015) in the R environment (R Core Team, 2018), because the task responses were binary ('acceptable' vs. 'unacceptable'). We set two fixed effects: (1) Time reference (*Past* vs. *Future*) and (2) Marker (*Na* and *Uri*). We also entered

https://doi.org/10.1017/langcog.2023.49 Published online by Cambridge University Press

two random effects in the model: Participant and Item. Before the computation of the models, we set the sum contrasts of the fixed effects as advised by Brehm and Alday (2022). Following Barr et al.'s (2013) recommendations, we started with a maximal model which included the random slopes for the fixed effects. We simplified the model until the convergence did not fail and until the fit was not singular as follows (Barr et al., 2013, p. 276; Myers, 2022). First, we used the maximal model; if it failed to converge, we computed the maximal model including the random slopes of the fixed effects but without random correlations; if it also failed to converge, the model only included random intercepts. We calculated the *p*-values of the model using the 'lmerTest' package (Kuznetsova et al., 2017). We conducted follow-up pairwise comparisons (*z*-scores) when interactions reached significance using the 'emmeans' package (Lenth, 2020), and we applied Bonferroni corrections on the *p* value of the post-hoc comparisons. The responses from the acceptability judgment task confirmed the pattern that we first assumed, in that *na*= after *katiaw* is acceptable, while it is not after *nutiaw*, and *uri*= is acceptable after *nutiaw* but not after *katiaw*. Therefore, we also considered accuracy rates of the judgment data.[9] We conducted the same statistical tests and procedure on the accuracy rates of the judgments as for the acceptability judgment tasks. We applied linear mixed-effect models on the log-transformed reaction time of the trials which the participants correctly judged.[10] We entered the factors of Time reference and Marker as fixed effects, and Participant and Item as random effects. As for the post-hoc comparisons, we performed follow-up paired *t*-tests, and the *p*-values were Bonferroni corrected.[11]

5. Results

5.1. Acceptability judgment and accuracy rates

The acceptability judgment ratings of the experimental sentences were as follows: 85.63% (SD = 35.10) of the Past-Na were acceptable, while this was the case for only 25.34% (SD = 43.53) of the Future-Na sentences. A reverse tendency was observed for *uri*=: 81.04% (SD = 38.22) of the sentences with *uri*= were acceptable after a deictic future time adverb, while the acceptance rate was of 21.14% (SD = 40.86) after a deictic past time adverb. See Fig. 1 for a graphical representation of the results.

We further verified these observations with statistical tests (see Table 4 for a summary of the statistics). The logistic mixed-effect models revealed no significant

[9]We computed the accuracy results as follows. For the grammatical sentences (i.e., Past-Na and Future-Uri), the accuracy results were the same as the acceptability judgment results. For the ungrammatical sentences (i.e., Future-Na and Past-Uri), we annotated trials judged as acceptable as inaccurately judged (switching '1' of the judgment data to '0' in the accuracy data).

[10]We also conducted the analysis on the reaction time of the all the trials, that is, the dataset also containing misjudged items, as well as on the dataset containing the outliers and not-log-transformed data (this alternative analysis plan was also adopted by Ono et al. (2020) in their sentence acceptability study on another Formosan language, Truku Seediq). The results were the same, so we only report the results for the analyses conducted on the trimmed and log-transformed data.

[11]Alternatively, when the Time reference × Marker interaction reached significance, we also ran analysis on a subset of the data (i.e., subset with only *na*=, another with only *uri*=). We made the choice to only report the results obtained with follow-up pairwise comparisons as they did not differ from the split-model method.

https://doi.org/10.1017/langcog.2023.49 Published online by Cambridge University Press

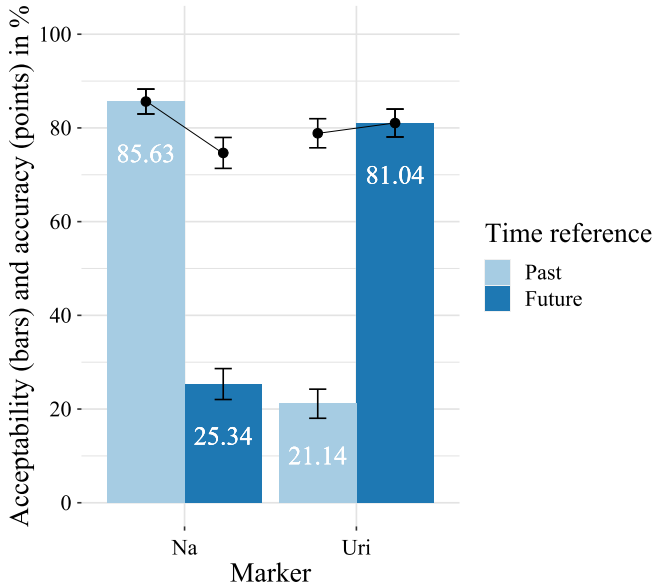

Figure 1. Percentage of the mean acceptability judgment of the sentences (bar charts) and the accuracy of the judgments (points) in terms of time reference and marker (error bars: 95% confidence interval).

Table 4. Summary of the statistical analysis on the acceptability judgment

	Estimate	*SE*	*z*-value	*p*-value
(Intercept)	0.2277	0.1310	1.739	0.08
Time reference	−0.0293	0.0514	−0.570	0.57
Marker	0.1700	0.0517	3.290	<.01
Time reference × Marker	−1.5668	0.0579	−27.064	<.001

*Model: glmer(Judgment ~ TimeRef*Marker + (1|subject) + (1|TrialID), family = binomial).*

main effect of Time reference ($\beta = -0.03$, SE = 0.05, $z(2683) = -0.57$, $p = .52$). The main effect of Marker was significant ($\beta = 0.17$, SE = 0.05, $z(2683) = 3.29$, $p < .01$), as well as the Time reference × Marker interaction ($\beta = -1.57$, SE = 0.06, $z(2683) = -27.06$, $p < .001$). Follow-up pairwise comparisons were directly conducted on the interaction. The contrast 'Past-Na versus Future-Na′ was significant ($z = -20.302$, $p < .0001$), as well as the one between Future-Uri and Past-Uri ($z = 20.147$, $p < .0001$). The contrast between the acceptable conditions (i.e., Past-Na and Future-Uri) was only marginally significant ($z = 2.585$, $p = 0.06$), and the contrast between the rejected conditions did not reach significance (i.e., Future-Na and Past-Uri; $z = 2.053$, $p = 0.24$).

Overall, the results from the judgment acceptability task show the following pattern: *na*= is generally acceptable after a past time adverb (i.e., *katiaw* 'yesterday'), but not after a future time adverb (i.e., *nutiaw* 'tomorrow'). We observed the reverse for *uri*=, which is generally acceptable after a future time adverb, but not after a past time adverb.

The mean accuracy rates of the acceptability judgments across the conditions were: Past-Na: 85.63% (SD = 35.10); Future-Na: 74.66% (SD = 43.53); Future-Uri:

https://doi.org/10.1017/langcog.2023.49 Published online by Cambridge University Press

Table 5. Summary of the statistical analysis on the accuracy rates of the acceptability data

	Estimate	SE	*z*-value	*p*-value
(Intercept)	1.5522	0.1753	8.856	<.001
Time reference	−0.1536	0.0510	−3.012	<.01
Marker	0.0208	0.0510	0.408	0.68
Time reference × Marker	−0.2234	0.0510	−4.379	<.001

*Model: glmer(correct ~ TimeRef*Marker + (1|subject) + (1|TrialID), family = binomial).*

81.04% (SD = 38.22); Past-Uri: 78.86% (SD = 40.86).[12] We compared the mean accuracy rates of the two markers with their respective baseline conditions instead of directly comparing the mean accuracy rates of *na*= and *uri*= since factors unrelated to the research question of the present study may have led to differences in the accuracy rates of the judgments of the two markers (e.g., different frequency of use of *na*= and *uri*=, etc.). For *na*=, sentences with a deictic past time adverb play the role of baseline condition to which sentences with a deictic future time adverb are compared. As for *uri*=, sentences with a deictic future time adverb are the baseline. At first sight, there seem to be greater differences in terms of mean accuracy of the judgments concerning *na*=. Past-Na sentences are judged more accurately than Future-Na sentences. A different pattern emerges for *uri*=. Despite the numeric differences between future and past time sentences, the mean accuracy does not seem to differ a lot.

We further verified these observations with inferential statistics (see Table 5). The logistic mixed-effect models revealed a significant main effect of Time reference ($\beta = -0.15$, SE = 0.05, $z(2683) = -3.01$, $p < .01$), but not of Marker ($\beta = 0.02$, SE = 0.05, $z(2683) = 0.41$, $p = .68$). The Time reference × Marker interaction also reached significance ($\beta = -0.22$, SE = 0.05, $z(2683) = -4.38$, $p < .001$). We further explored the interaction with follow-up pairwise comparisons. The contrast Past-Na versus Future-Na was highly significant, with Past-Na judged more accurately than Future-Na ($z = -5.148$, $p < 0.001$). Contrastively, the comparison Future-Uri versus Past-Uri did not reach significance ($z = 0.983$, $p = 1$).

Overall, the data from the mean accuracy rates of the acceptability judgments reveal a different pattern than the data from the acceptability judgment. The difference between Past-Na and Future-Na suggests that judging *na*= in a temporally speaking incoherent context is more difficult that in a coherent temporal context. On the contrary, we found no significant differences concerning *uri*= in a future and past time context in terms of accuracy rate, despite the obvious difference in terms of acceptability judgment.

5.2. Reaction time

The mean reaction time data per condition is as follows: Past-Na: 4742 ms (SD = 1072); Future-Na: 5010 ms (SD = 1705); Future-Uri: 4761 ms (SD = 1271); Past-Uri: 4829 ms (SD = 1504). These data are graphically represented in Fig. 2.

[12]A response was 'accurate' when the participants accepted grammatical sentences or rejected ungrammatical sentences. In other words, we coded 'rejected grammatical sentences' and 'accepted ungrammatical sentences' as inaccurate responses. See Footnote 6 and Section 4.3 for the justification of these considerations.

https://doi.org/10.1017/langcog.2023.49 Published online by Cambridge University Press

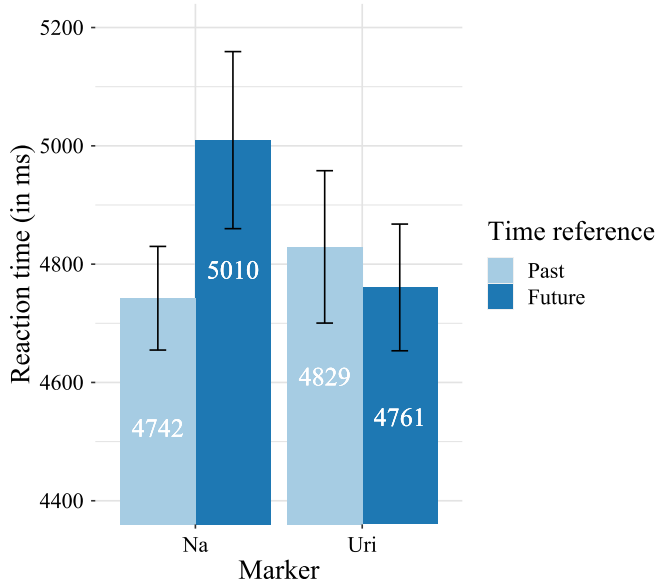

Figure 2. Mean reaction time of the sentences (in ms) in terms of time reference and marker (error bars: 95% confidence interval).

The participants judged Past-Na sentences faster than Future-Na sentences, and Future-Uri sentences faster than Past-Uri sentences. Despite the similarity of the patterns, they are not exactly identical: the difference in reaction time for *na*= is numerically greater for *uri*=.

We further assessed these observations with statistical analyses (see Table 6). Linear mixed-effect models revealed no significant main effect of Time reference or Marker (Time reference: β = 0.006, SE = 0.004, $t(2148) = 1.55$, $p = .12$; Marker: β = 0.006, SE = 0.004, $t(2148) = 1.47$, $p = .14$). The Time reference x Marker interaction was significant (β = 0.013, SE = 0.004, $t(2148) = 3.20$, $p < .01$). Follow-up pairwise t-tests on the interaction showed a significant contrast between Past-Na and Future-Na ($t = 3.363$, $p < .01$), with a longer reaction time value for Future-Na. The contrast Future-Uri versus Past-Uri did not reach significance ($t = -1.162$, $p = 1$).

In sum, the data from the reaction time also display a difference between the temporal processing of *na*= and the temporal processing of *uri*=.

Table 6. Summary of the statistical analysis on the reaction time

	Estimate	SE	*t*-value	*p*-value
(Intercept)	8.47	0.04	195.81	<.0001
Time reference	0.006	0.004	1.55	0.12
Marker	0.006	0.004	1.47	0.14
Time reference × Marker	0.013	0.004	3.195	<.01

*Model: lmer(log(latency) ~ TimeRef*Marker + (1|subject) + (1|TrialID), family = binomial).*

https://doi.org/10.1017/langcog.2023.49 Published online by Cambridge University Press

6. Discussion

6.1. Past time reference processing decoupled from future time reference in Paiwan

Selecting Paiwan as the language of research allowed us to explore how Paiwan speakers compute temporal relations between deictic time adverbs and grammatical morphemes referring to perfective aspect (use constrained in a realis mood context) and irrealis mood. To this aim, the present study investigated how the temporal concord violation of the two markers *na=* and *uri=* in Paiwan are processed by placing them either after a deictic past or future time adverb. We stated two hypotheses. The first account hypothesized that processing the temporal concord violation of the past-time related grammatical marker would be more effortful than the one of the non-past-time related grammatical marker (in line with the PADILIH; Bastiaanse, 2018; Dragoy et al., 2012). This hypothesis predicts that *na=* in a temporal concord violation configuration is harder to process than *uri=*. The second account drew on the idea that irrealis is harder to process than realis because it involves the mental representation of an event that is only imagined or hypothesized (Lyons, 1977; Mucha, 2015; Smith & Erbaugh, 2005; Suddendorf, 2010; Suddendorf et al., 2009). Therefore, the judgment of the irrealis marker *uri=* should be harder to process when compared with its baseline than the realis marker *na=* under this hypothesis.

In our experiment, we found differences between the processing of *na=* and *uri=* in a temporal concord violation context concerning the accuracy rates of the acceptability judgments: the temporal concord violation of *na=* was harder to process (when compared with its grammatical baseline), as reflected by the lower accuracy rate of the Future-Na condition. However, no significant difference was found between *uri=* placed after a deictic future time adverb or a deictic past time adverb, suggesting that the temporal concord violation of *uri=* was not more effortful than its grammatical counterpart. The reaction time results also exhibit a different pattern between *na=* and *uri=*. Judging ungrammatical sentences with *na=* required more time while no significant difference was found concerning sentences with *uri=*. These again suggest that processing *na=* is more effortful than *uri=* .

Our results clearly support the PADILIH, which states that processing past-time-related markers is harder than non-past-time-related markers. Indeed, when compared with their respective baseline conditions, the temporal concord violation of *na=* (past-time related marker) led to significantly lower accuracy rates of the judgments as well as a significantly longer reaction time, while such a difference was not significant concerning *uri=* (non-past-time related marker).

We can interpret the asymmetry of the processing of the temporal concord of *na=* and *uri=* as the increased difficulty of processing the temporal concord violation of past time reference when compared with its grammatical baseline, which is not found for non-past time reference in general. Indeed, the results we obtained in the present study are similar to the ones reported by Dragoy et al. (2012), where they found that when judging the temporal concord violation of Dutch sentences (verb inflected either for past tense or present tense), the violation was detected later when the verb had a past tense form (reflected by a longer reaction time), and less accurately (when compared with their respective baselines). Biondo et al. (2022) report similar results for Spanish. This similarity of pattern despite the differences between Paiwan, Dutch, and Spanish and the experimental designs (time reference expressed with optional mood morphemes in Paiwan and with obligatory tense morphemes in Dutch and Spanish) suggests that the processing of time reference relies on shared

https://doi.org/10.1017/langcog.2023.49 Published online by Cambridge University Press

cognitive mechanisms across languages. According to the PADILIH, such cognitive mechanisms are related to the retrieval of the reference time when computing the meaning of the sentences. Namely, when the verb is inflected with a non-past-time-related marker (i.e., *uri=* in Paiwan, present tense morpheme in Dutch or future tense morpheme in Spanish), the reference time aligns with the time of speech, such that the reference time does not need to be retrieved from the previous context.[13] In other words, the reference time is provided (and therefore the temporal interpretation of the sentence) as soon as the verb inflected for a non-past-time related morpheme appears in the sentence, and the participants are directly able to compute the integration of the temporal verbal information with the temporal information given previously by the deictic time adverb. The PADILIH assumes different processing mechanisms when the verb is inflected by a past-time-related morpheme (*na=* in Paiwan or past tense morpheme in Dutch or in Spanish): the detection of the violation of past-time-related morphemes is delayed and less accurate.

6.2. The interaction of perfective aspect when processing time reference during language comprehension

An alternative account of the underlying mechanisms hypothesized by the PADILIH can be proposed. Indeed, *na=* and *uri=* appears not only to minimally differ based on the realis status they express (*na=* for realis mood, *uri=* for irrealis mood), but also according to the aspectual meaning they involve. As presented in Section 2.3, while *uri=* is quite neutral regarding such aspectual distinctions, *na=* is marked for perfective aspect. Therefore, it may also be the case that the additional processing cost of the temporal concord violation of *na=* is not due to the processing of past time reference at the cognitive level, but alternatively due to the temporal concord violation of perfective aspect.

Several pieces of evidence converge this hypothesis. Dragoy et al. (2012) reported a difference of processing between Dutch past tense and present tense verbs: participants judged past tense verbs less accurately and more slowly than present tense verbs. The picture may be more complex. As Boogaart (1999, pp. 131–166) observes, the Dutch *simple past* (i.e., the verbal form used in the past condition in Dragoy et al., 2012), unlike its label, does not simply locate an event in the past but also expresses perfective aspect meaning. Therefore, the alternative explanation proposed here also holds for the results found in Dragoy et al. (2012): it may also be the case that the lower accuracy and slower RTs found for past tense verbs are also due to the perfective meaning conveyed by these verbs.

This also corroborates results found in other languages, including the Formosan languages. Tang (2019) conducted a production study (story telling based on 'the Frog Story') in Truku Seediq with speakers of different age ranges (older adults, adults, young adults, and youth), which corresponded to different levels of proficiency in Truku Seediq (namely, the older the speakers, the more proficient they are in Truku Seediq). The production results indicate that the younger/the less proficient the speakers are, the more likely they drop the perfective morpheme while telling the story. By contrast, other verbal morphemes were relatively well-preserved, including

[13]Concerning *uri=*, the PADILIH claims that present and future time reference share similar properties in terms of the retrieval of the reference time, since future time can be seen as the projection of the present time.

https://doi.org/10.1017/langcog.2023.49 Published online by Cambridge University Press

voice marking such as AV marking, and, crucially for the present discussion, other aspectual marking such as progressive aspect. This suggests that perfective aspect in particular induces greater difficulties, as also reported in other acceptability judgment studies (e.g., in Mandarin Chinese; Collart & Chan, 2021). These results may indicate that the additional processing costs reported in Dragoy et al. (2012) and in the present study are not only due to the processing of past time reference *per se*, but also due to the perfective meaning conveyed by the verbal forms.

What does make perfective aspect harder to process when placed in a temporal concord violation configuration? One explanation draws on the temporal relation induced by perfective aspect regarding the time of speech. Perfective aspect entails that the event is seen as a whole. In the above-mentioned experiments and in the present study, the temporal interpretation of the perfective verbs was forced by placing them after deictic time adverbs, such that the participants had to interpret this whole event as being either in the past or in the future. Because of this characteristic, perfective aspect is prototypically related to past time reference. However, unlike past tense, perfective aspect markers only provide a specific viewpoint to the event denoted by the verb, i.e., the event is 'ready' for other operations from the perspective of the time of speech, such as the view of the event as in a sequence related to the time of speech rather than a simple temporal localization (Baggio et al., 2012; Collart, 2022). Such a conceptualization of a 'whole event' is less compatible with future time reference, reflected by a harder (even impossible in some languages) integration in a future context. In other words, it is harder for the participants to retrieve the reference time of the sentence when a perfective aspect marker occurs with a deictic future time adverb, hence the lower accuracy rate and longer reaction time. This explanation is in line with the results of Dragoy and Bastiaanse (2013) in Russian, where they found that perfective aspect and tense marking interact when it comes to the production of time reference marking by agrammatic participants. The prototypical cognitive bridge between perfective aspect and past time reference briefly exposed here is better preserved when coupled with past tense. In sum, this alternative explanation does not go against the PADILIH itself since this explanation is also related with the conceptual/prototypical link between perfective aspect marking and past time reference, but it suggests underlying mechanisms which are slightly different from the ones originally proposed by the PADILIH in Bastiaanse et al. (2011).

6.3. Contributions and limitations of the present study

This study reports behavioral results regarding the processing of time reference using the temporal concord paradigm in Paiwan, a Formosan language exhibiting realis and irrealis distinction with clitics whose one of them (*na*=, realis) closely interacts with perfective aspect meaning. As a result, it brings new insights to this growing topic in the psycholinguistic and neurolinguistic literature by offering results from a Formosan language involving grammatical marking rarely – if not – found in Indo-European languages. Despite the original contribution that this enterprise provides, this is not without limitations. First, because of historical and sociolinguistic factors, the age of native speakers of Formosan languages like Paiwan with high proficiency is relatively high, i.e., generally higher than 40 or 50 years old. Younger generations do not exhibit the same level of proficiency, or only low proficiency in

https://doi.org/10.1017/langcog.2023.49 Published online by Cambridge University Press

Formosan languages (see Tang, 2021 for a detailed survey on this issue on Truku Seediq). Crucially, the participants in other experiments cited in this study (Collart & Chan, 2021; Dragoy et al., 2012) are much younger than the participants recruited in the present study. Therefore, although these two studies and the present one report similar result patterns, one cannot exclude that the data described here also involve additional factors such as the aging process.

A second limitation of this study, also due to historical and sociolinguistic factors, is that all the participants recruited in the present study are bilingual. Even if their native language is Paiwan, they are all able to communicate at least in Mandarin Chinese (Tang, 2021). Therefore, we cannot discard the possibility that their knowledge of Mandarin Chinese may have influenced the processing of *na=* and *uri=* in temporal concord contexts. This idea further corroborates the fact that the results found for *na=* in the present study are similar to the ones found for *-le* in Mandarin Chinese (Collart & Chan, 2021). But some indicators suggest that this limitation may not have influenced the data so much. First, Paiwan is still the dominant language of the participants recruited in the present study. Second, Dragoy et al. (2012) found the same processing pattern in Dutch. Therefore, this rather indicates that this processing pattern is stable across speakers of different languages.

Finally, the sample size of our study ($N = 18$) is not large. This is due to the strict criteria we set when recruiting participants, notably the fact that we recruited participants who grew up and lived most of the time in the same village. Participants from other villages could be recruited to compensate for this sample size in future studies.

7. Conclusion

This study explores the issue of the processing of temporal concord using a sentence acceptability judgment task. The Past Discourse Linking Hypothesis, supported by experimental results from Dutch and English (tense languages), states that detecting the temporal concord violation between a non-past time adverb and verb expressing past time reference is more difficult than a sentence with a past time adverb and a verb expressing present or future time reference. In this study, we tested the PADILIH with Paiwan (Formosan language), where time reference can be expressed with realis mood interacting with perfective aspect, and irrealis mood. Despite these grammatical and semantic differences, the results of the experiment are in line with previous related studies and support the PADILIH. After a closer inspection, it appeared that the past-time-reference disadvantage reported in previous experiments as well as in this paper crosses with perfective aspect marking, leaving the question of the effect of aspectual marking on the computation of temporal relations when tense and realis are involved during online processing open. Future research targeting other languages is needed to further explore this issue. In the future, it would be interesting to compare the data collected for this dialect with experiments done in other regions, as dialectal variation or degrees of comprehension might exist.

On a final note, the present study can also be taken as an instance of how the domains of field linguistics and experimental linguistics can work jointly and learn from each other, adding another case to a growing body of literature and more insights from a methodological point of view.

https://doi.org/10.1017/langcog.2023.49 Published online by Cambridge University Press

Acknowledgement. We wish to express our gratitude to Drungdrung Tjakisuvung and Chia-Hsing Rick Sung for their assistance during the material creation and data collection process. We also thank the audience of the *Architectures and Mechanisms for Language Processing 28* as well as Shiaohui Chan for their comments on previous draft of this study. Last, but not least, we are grateful to the reviewers for their very constructive comments.

Data availability statement. The experimental data, including the experimental sentences, the anonymized raw results, and R script used for the analysis of this study can be found in the following OSF repository: https://osf.io/q35sz/.

Financial disclosure. This work was funded by the Institute of Linguistics of Academia Sinica in 2020.

Competing interest. The authors declare none.

References

Anand, P., Chung, S., & Wagers, M. (2011). *Widening the net: Challenges for gathering linguistic data in the digital age*. White paper published in NSF project SBE 2020: Future research in the Social, Behavioral and Economics Sciences.

Baggio, G. (2008). Processing temporal constraints: An ERP study. *Language Learning*, 58(s1), 35–55.

Baggio, G., van Lambalgen, M., & Hagoort, P. (2012). Language, linguistics and cognition. In R. Kempson, T. Fernando, & N. Asher (Eds.), *Handbook of the philosophy of science* (Vol. 14, pp. 325–355). North Holland.

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.

Bastiaanse, R. (2018). What language impairments can tell us about the organization of language in the brain. In B. Sharp, W. Lubaszewski, & F. Sedes (Eds.), *Natural language processing and cognitive science – Proceedings 2018* (pp. 9–18). Jagiellonian Library.

Bastiaanse, R., Bamyaci, E., Hsu, C., Lee, J., Yarbay Duman, T., & Thompson, C. K. (2011). Time reference in agrammatic aphasia: A cross-linguistic study. *Journal of Neurolinguistics*, 24(6), 652–673.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48.

Bhat, D. N. S. (1999). *The prominence of tense, aspect and mood*. John Benjamins.

Biondo, N., Bergamini, E., & Vespignani, F. (2021). The processing of temporal concord in sentences: An ERP investigation into the role of verb-adverb distance. In V. Torrens (Ed.), *Syntax processing* (pp. 1–47). Cambridge Scholars Publishing.

Biondo, N., Soilemezidi, M., & Mancini, S. (2022). Yesterday is history, tomorrow is a mystery: An eye-tracking investigation of the processing of past and future time reference during sentence reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 48(7), 1001–1018.

Boogaart, R. J. U. (1999). *Aspect and temporal ordering: A contrastive analysis of Dutch and English*. Holland Academic Graphics.

Bos, L. S., Dragoy, O., Stowe, L. A., & Bastiaanse, R. (2013). Time reference teased apart from tense: Thinking beyond the present. *Journal of Neurolinguistics*, 26(2), 283–297.

Brehm, L., & Alday, P. M. (2022). Contrast coding choices in a decade of mixed models. *Journal of Memory and Language*, 125, Article 104334.

Chang, H. A. (2006). *A reference grammar of Paiwan* [Doctoral dissertation, Australian National University]. Australian National University Open Access Theses. https://openresearch-repository.anu.edu.au/handle/1885/10719

Collart, A. (2022). *Bridging time in language and in the brain: Insights from Mandarin Chinese* [Doctoral dissertation, National Taiwan Normal University]. National Digital Library of Theses and Dissertations in Taiwan. https://hdl.handle.net/11296/w7mrm5

Collart, A. (2023). Experimental linguistics embracing linguistic diversity: On the contributions of Formosan languages to models of sentence processing. In P. Li, E. Zeitoun, & R. D. Busser (Eds.), *A handbook of Formosan languages: The indigenous languages of Taiwan, part 2* (pp. 22–61). Brill.

Collart, A., & Chan, S. (2021). Processing past time reference in a tenseless language: An ERP study on the Mandarin aspectual morphemes *-le* and *-guo*. *Journal of Neurolinguistics*, 59, Article 100998.

https://doi.org/10.1017/langcog.2023.49 Published online by Cambridge University Press

Comrie, B. (1976). *Aspect: An introduction to the study of verbal aspect and related problems*. Cambridge University Press.

Comrie, B. (1985). *Tense*. Cambridge University Press.

Council of Indigenous People. (2023). *Yuanzhuminzu renkoushu tongji ziliao* [Census of indigenous population]. https://www.cip.gov.tw/zh-tw/chart/data-list/15805EC819F939F5/index.html?cumid=D0636733C6861689&pid=9A0831149DC993FA (Consulted on September 20, 2023).

Crowley, T. (2007). *Field linguistics: A beginner's guide*. Oxford University Press.

Dahl, Ö. (1985). *Tense and aspect systems*. Blackwell.

Dahl, Ö. (2015). How WEIRD are WALS languages? In *Conference presentation at the diversity linguistics: Retrospect and prospect conference, Leipzig, Germany*.

de Haan, F. (2012). Irrealis: Fact or fiction? *Language Sciences*, 34(2), 107–130.

Dragoy, O., & Bastiaanse, R. (2013). Aspects of time: Time reference and aspect production in Russian aphasic speakers. *Journal of Neurolinguistics*, 26(1), 113–128.

Dragoy, O., Stowe, L. A., Bos, L. S., & Bastiaanse, R. (2012). From time to time: Processing time reference violations in Dutch. *Journal of Memory and Language*, 66(1), 307–325.

Egli, H. (1989). *Mirimiringan. Die mythen und märchen der Paiwan. Das frühe weltbild des Fernen Ostans* [Mirimiringan: Myths and folktales of the Paiwan. The early view of the Far East.] Zurich: Verlag Die Waage [In German]

Elliott, J. R. (2000). Realis and irrealis: Forms and concepts of the grammaticalization of reality. *Lingusitic Typology*, 4, 55–90.

Faroqi-Shah, Y., & Dickey, M. W. (2009). On-line processing of tense and temporality in agrammatic aphasia. *Brain and Language*, 108(2), 97–111.

Ferrell, R. (1982). *Paiwan dictionary*. Australian National University.

Ferrell, R., & Tjakisuvung, D. (2023). Paiwan texts. In E. Zeitoun, & C. Chang (Eds.) [Manuscript in preparation]. Institute of Linguistics, Academia Sinica.

Haspelmath, M. (2001). The European linguistic area: Standard average European. In M. Haspelmath, E. König, W. Oesterreicher, & W. Raible (Eds.), *Language typology and language universals: An international handbook* (pp. 1492–1510). Mouton de Gruyter.

Heinrich, J., Heine, S. J., & Norenzayan, A. (2010). Most people are not WEIRD. *Nature*, 466(1), 29.

Huang, W. (2012). *A study of verbal morphology in Puljetji Paiwan* [Master's thesis, National Tsing Hua University]. National Digital Library of Theses and Dissertations in Taiwan. https://hdl.handle.net/11296/3u5vc2

Kidd, E., & Garcia, R. (2022). How diverse is child language acquisition? First Language (Special issue: How diverse is child language acquisition research? *Sage Journal*, 42(6), 703–735.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 1–26.

Lenth, R. (2020). *emmeans: Estimated Marginal Means, aka Least-Squares Means (version 1.4.6).* R Project. https://cran.r-project.org/packages/emmeans/index.html

Li, C. N., & Thompson, S. A. (1981). *Mandarin Chinese: A functional reference grammar*. University of California Press.

Li, W. (2012). Temporal and aspectual references in Mandarin Chinese. *Journal of Pragmatics*, 44(14), 2045–2066.

Lyons, J. (1977). 17 modality. In *Semantics* (Vol. 2, pp. 787–849). Cambridge University Press.

Maia, M. (2021). Non-WEIRD experimental field work as *bricolage*: A discourse on methods in the investigation of deixis and coreference in the Karajá language of Central Brazil. *Journal of Cultural Cognitive Science*, 5, 101–112.

Majid, A., & Levinson, S. C. (2010). WEIRD languages have misled us, too. *Behavior and Brain Sciences*, 33 (2–3), 103.

Mucha, A. (2015). Temporal interpretation and cross–linguistic variation. [Doctoral dissertation, University of Potsdam]. URN:NBN Resolver for Germany & Switzerland. http://nbnresolving.de/urn:nbn:de:kobv:517-opus4-85935

Myers, J. (2022). *Yet another statistics-for-linguistics book* [Unpublished manuscript]. Graduate Institute of Linguistics, National Chung Cheng University.

Ng, I.-s. (a.k.a. Huang, W.). (2023). Paiwan. In P. Li, E. Zeitoun & R. De Busser (Eds.), *A handbook of Formosan languages: The indigenous languages of Taiwan, part 3* (pp. 294–328). Brill.

https://doi.org/10.1017/langcog.2023.49 Published online by Cambridge University Press

Ono, H., Kim, J., Sato, M., Tang, A. A.-Y., & Koizumi, M. (2020). Syntax and processing in Seediq: A behavioral study. *Journal of East Asian Linguistics*, 29(2), 237–258.
Palmer, F. R. (2001). *Mood and modality*. Cambridge University Press.
Partee, B. H. (1973). Some structural analogies between tenses and pronouns in English. *Journal of Philosophy*, 70(18), 601–609.
Plungian, V. (2005). Irrealis and modality in Russian and in typological perspective. In B. Hansen & P. Karlik (Eds.), *Modality in Slavonic languages: New perspectives* (pp. 135–146). Sagner Verlag.
R Core Team. (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/
Reichenbach, H. (2005). The tenses of verbs. In I. Mani, J. Pustejovsky and R. Gaizauskas (Eds.), *The language of time: A reader* (pp. 71–78). Oxford University Press (Original work: Reichenbach, H. (1947). *Elements of symbolic logic.* Dover Publications.)
Ross, M. D. (1995). Reconstructing Proto-Austronesian verbal morphology: Evidence from Taiwan. In P. Li, C. Tsang, Y. Huang, D. Ho, & C. Tseng (Eds.), *Austronesian studies relating to Taiwan* (pp. 727–791). Institute of History and Philology, Academia Sinica.
Smith, C. S., & Erbaugh, M. S. (2005). Temporal interpretation in Mandarin Chinese. *Linguistics*, 43(4), 713–756.
Speed, L. J., Wnuk, E., & Majid, A. (2018). Studying psycholinguistics out of the lab. In A. M. B. de Groot & P. Hagoort (Eds.), *Research methods in psycholinguistics and the neurobiology of language: A practical guide* (pp. 190–207). Blackwell.
Suddendorf, T. (2010). Episodic memory versus episodic foresight: Similarities and differences. *WIREs Cognitive Science*, 1(1), 99–107.
Suddendorf, T., Addis, D. R., & Corballis, M. C. (2009). Mental time travel and the shaping of the human mind. *Philosophical Transactions of the Royal Society B*, 364, 1317–1324.
Sung, C. (2006). *Temporal expressions in Paiwan* [Master's thesis, National Chung Cheng University]. National Digital Library of Theses and Dissertations in Taiwan. https://hdl.handle.net/11296/267y73
Tang, A. A. (2019). Linguistic features of attrition in Truku Seediq. *Taiwan Indigenous Studies Review*, 25, 1–40.
Tang, A. A. (2021). *From diagnosis to remedial plan: A psycholinguistic assessment of language shift, intergenerational linguistic proficiency, and language planning in Truku*. National Dong Hwa University, College of Indigenous Studies.
von Prince, K., Krajinović, A., & Krifka, M. (2022). Irrealis is real. *Language*, 98(2), 221–249.
Wagers, M., & Chung, S. (2023). Language processing experiments in the field. In J. Sprouse (Ed.), *The Oxford handbook of experimental syntax* (pp. 491–511). Oxford University Press.
Whalen, D. H., & McDonough, J. (2015). Taking the laboratory into the field. *Annual Review of Linguistics*, 1, 395–415.
Yano, M., Niikuni, K., Ono, H., Sato, M., Tang, A. A.-Y., & Koizumu, M. (2019). Syntax and processing in Seediq: An event-related potential study. *Journal of East Asian Linguistics*, 28(4), 395–419.
Zagona, K. (2003). Tense and anaphora: Is there a tense-specific theory of coference? In A. Barss (Ed.), *Anaphora: A reference guide* (pp. 140 –171). Blackwell.
Zeitoun, E., Huang, L. M., Yeh, M. M., Chang, A. H., & Wu, J. J. (1996). The temporal/aspectual and modal systems of the Formosan languages: A typological perspective. *Oceanic Linguistics*, 35(1), 21–56.

Appendix A. Notes on the methodology of psycholinguistic studies on endangered languages

The present study involves the use of experimental techniques to investigate the sentence processing of an endangered language. Some notes are necessary regarding some characteristics of experimental linguistic studies conducted in the field on endangered languages in order to understand the methodological choices of the present study, notably concerning (1) the recruitment and pool of participants, (2) the crucial involvement of a native speaker consultant, (3) the familiarity of the task administrated during the experiment, and (4) the importance of the interaction and debriefing with the participants. We briefly discuss each of these below.

Recruiting the participants of endangered languages turns out to be much more difficult than for languages with a lot of speakers for a number of reasons discussed in Crowley (2007) and Speed et al.

https://doi.org/10.1017/langcog.2023.49 Published online by Cambridge University Press

(2018). Concerning the present study, as we could not be completely sure that speakers of different dialects (located in different villages) may exhibit the same linguistic behavior as in the village where our observations were made, we restricted the pool of participants to only one village. While this could ensure more hegemony among the participants, the potential number of participants was consequently very low. Since this was predictable, the number of sentences in the present study is much higher than in typical psycholinguistic studies in order to compensate for the lower number of participants (Speed et al., 2018, p. 194). We created all the sentences with native speakers, and notably with one native consultant who received a linguistic training in the past. This collaboration appears also to be typical of such psycholinguistic studies (Wagers & Chung, 2023).

The background of the participants also greatly differs from many psycholinguistic studies, and so may be the results. Indeed, the participants are often older and not used to experimental settings, even for tasks including acceptability judgments, when compared with typical participants such as university students (Heinrich et al., 2010). As Speed et al. (2018, p. 194) mention, "things that seem unproblematic from the point of view of university students, who spend hours listening to lectures and writing exams on a daily basis (e.g., performing repetitive tasks), can be highly demanding for other people". How may this observation affect the results of psycholinguistic experiments? This may surely introduce more variability in the performance of the task, as well as result patterns that are less extreme or marked than usual. Collart (2023) points out that it is common in psycholinguist studies of Formosan languages that for comprehension questions, accuracy rates are in the range of 75 to 80%, when it is usually above 95% concerning university students (Ono et al., 2020; Tang, 2021). This range can also be down below 75% for tasks involving judging match and mismatch conditions (Yano et al., 2019).

There are several ways to compensate for this problem. First, the involvement of a native speaker trained in linguistics, in addition to the collaboration of linguists, is crucial to have a comprehensive analysis of the phenomenon under investigation (Whalen & McDonough, 2015, pp. 8, 12). A second way is to rely more than usual on the interaction with the participants by debriefing the experiments and obtain their feedback on the nature and the procedure of the experiments, as well as their own thoughts concerning the research question and the stimuli (Wagers & Chung, 2023).

Cite this article: Collart, A., & Zeitoun, E. (2023). Past and future time reference processing teased apart in Paiwan, an endangered Formosan language, *Language and Cognition*, 1–26. https://doi.org/10.1017/langcog.2023.49