Experimental research in cross-linguistic psycholinguistics

Sebastian Sauppe[1,2], Caroline Andrews[1,2], and Elisabeth Norcliffe[3]

[1]*Department of Comparative Language Science, University of Zurich*

[2]*Center for the Interdisciplinary Study of Language Evolution, University of Zurich*

[3]*Department of Experimental Psychology, University of Oxford*

1 INTRODUCTION

This chapter provides an overview of cross-linguistic psycholinguistics research, including work that is conducted outside of traditional lab contexts and/or in the field. The goal of the chapter is to provide, where possible, practical examples and advice for researchers who are interested in adapting current experimental psycholinguistic methods for a cross-linguistic context. While most of the examples and specifics focus on sentence processing research, many of the recommendations and discussion can be applied to other areas of experimental language research, such as experimental phonology.

2 HISTORICAL PERSPECTIVES ON CROSS-LINGUISTIC PSYCHOLINGUISTICS

In 1995, Inoue and Fodor wrote that it "is easier to argue that parsing Japanese is impossible than to explain how it is done." They pointed to three properties of Japanese

that should, theoretically, pose a problem for the human parser: (i) verb-finality, (ii) scrambling, and (iii) null pronouns/argument drop.

On the face of it, this *should* have been an odd thing to have to say. None of the three properties must pose a real problem for sentence comprehension, given that Japanese speakers do, of course, successfully understand their interlocutors. But, more to the point of psycholinguistic theory, none of these properties is especially rare cross-linguistically. Indeed, SOV is the most common basic word order among the world's languages (Dryer, 2013b). Languages which have the exact configuration of null subject pronouns seen in Japanese are perhaps somewhat unusual in typological samples ($\sim$ 9% in Dryer, 2013a), but adding in languages which allow null NPs in canonical position combined with marking on the verb covers 73% of languages in Dryer's sample, mostly because of the sheer pervasiveness of verbal agreement. Estimates of the typological prevalence of scrambling and object null pronouns are somewhat trickier to come by (Saleem, 2010; Sekerina, 1997), but neither are exceptionally rare and indeed may be more common than not.

But of course, only part of Inoue and Fodor's observation was about the objective difficulty of parsing Japanese; much of it was about the difficulty of accounting for Japanese parsing given the psycholinguistic theories *at that moment*. At the time, sentence processing had been almost exclusively focused on English, which critically has none of the three properties that so worried Inoue and Fodor. More than 25 years on, the study of both Japanese and the specific properties that Inoue and Fodor highlighted have been important lines of research in psycholinguistics, leading to advances not just in our understanding of parsing in a single language, but for aspects of psycholinguistic theory more generally. Verb-finality has provided a window into incrementality of both production (Hwang and Kaiser, 2014; Momma and Ferreira, 2019; Momma, Slevc, and Phillips, 2016; Sauppe, 2017b; Schriefers, Teruel, and Meinshausen, 1998) and comprehension (Konieczny, Hemforth, Scheepers, and Strube, 1997), as well as many other issues,

including filler-gap dependencies (Aoshima, Phillips, and Weinberg, 2004), prediction and locality (Kamide, Altmann, and Haywood, 2003; Levy and Keller, 2013; Mitsugi, 2017), and turn-taking (Barthel and Sauppe, 2019). Scrambling has been studied for its own sake (Nakano, Felser, and Clahsen, 2002; Sekerina, 2003) and also used as a tool to provide insight into many other phenomena, especially the delicate work of teasing apart how the parser relies on the often co-occurring cues of case, word order, and grammatical function (Tamaoka, Sakai, Kawahara, Miyaoka, Lim, and Koizumi, 2005; Yamashita, 1997) and information structure (Ferreira and Yoshita, 2003). Likewise, null pronouns have been explored in Romance and East Asian languages and have contributed greatly to studies of information structure (Runner and Ibarra, 2016), notions of accessibility in processing (Carminati, 2002), optionality (Saah, 1995) and beyond.

Japanese is but one of the more than 7000 languages spoken in the world today, each one offering 'a natural laboratory of variation' (Evans and Levinson, 2009) against which psycholinguistic theories can be tested and developed. The current empirical base of language processing research represents a small fraction of this linguistic diversity, still limited largely to Germanic and Romance families and a handful of also large East Asian languages (Blasi, Henrich, Adamou, Kemmerer, and Majid, 2022; Jaeger and Norcliffe, 2009; Norcliffe, Harris, and Jaeger, 2015a). There have been many calls to broaden the typological diversity in psycholinguistics (Anand, Chung, and Wagers, 2010; Bates, Devescovi, and Wulfeck, 2001; Chung, Borja, and Wagers, 2012; Clahsen, 2016; Cutler, 1985; Evans and Levinson, 2009; Hellwig, 2019; Jaeger and Norcliffe, 2009; MacWhinney and Bates, 1989a; Nielsen, Haun, Kärtner, and Legare, 2017; Norcliffe et al., 2015a, *inter alia*). Not surprisingly, the problem also exists in child language acquisition (Kidd and Garcia, 2022; Slobin, 2014; Stoll, 2009). In recent years, the cognitive sciences more generally have recognized the need for more diversity (Arnett, 2008), in particular the well-known call by Henrich, Heine, and Norenzayan (2010) to move beyond WEIRD (Western Educated Industrialized Rich Democratic) populations and English speaking

participants (see also Blasi et al., 2022). Within this, psycholinguistics is particularly well positioned to notice the opportunity to be had in looking beyond western(ized) labs, given the long-standing and productive research traditions in adjacent fields of linguistic typology, comparative linguistics and language documentation, which directly engage with linguistic variation and bring it into central focus.

Since Inoue and Fodor's article, the landscape of cross-linguistic psycholinguistics research has shifted considerably. There is now a growing group of researchers producing considerably more cross-linguistic work than ever before.[1]

3 CRITICAL ISSUES AND TOPICS

To the extent that psycholinguistic theory makes any claims about *human* cognition or how *human* language is processed—whether those claims are about the potential range of variation or characteristics that maybe universal—it is necessary to expand cross-linguistic coverage. Doing this serves two functions. First, replications of previously established effects in new languages provide critical empirical support for the (assumed) universality of processing mechanisms. Second, researching new languages makes available new linguistic phenomena that fall outside of the scope of current psycholinguistic theories. Thus, a productive approach is to select a "target" language that has grammatical properties that provide a new window on current psycholinguistic theories, for example, because existing theories make no or conflicting predictions about how certain features of the grammar should influence processing or because a language allows testing a claim that is not possible to examine with other languages.

Additionally, cross-linguistic research presents an opportunity to expand the field methodologically. Research in non-lab contexts, especially with technical equipment

[1] See `https://docs.google.com/spreadsheets/d/1AS6NFJad5pqg0gY9g8R8AgGEixTyx0KbGtJ-SLEGG84/edit?usp=sharing` for a running database of researchers working on languages underrepresented in psycholinguistic research. It is actively maintained at least as of December 2022, but inclusion is largely self-selected so it is likely to be incomplete.

such as eyetrackers or EEG, comes with challenges that researchers typically avoid in the lab. Until recently, many of these challenges were largely insurmountable. New technological advances have made methods much more portable, but new environments present a much greater range of lighting, electrical, social, and political issues, etc., all of which can impact data interpretation. Moreover, new populations themselves come with challenges that demand adaptation. For instance, many community languages are not written as frequently, and therefore speakers may not be literate in their native language, even when they are literate in the local *lingua franca*. Additionally, Western researchers are likely to underestimate the impact that constant experience with technology has had on their typical undergraduate subjects. This can mean that critical timing issues that were painstakingly established in the lab — for example the appropriate stimulus onset time in a picture-word interference study (e.g., Meyer, 1996) — may need to be re-discovered when working outside of the laboratory and with new populations.

4 CURRENT CONTRIBUTIONS AND RESEARCH

In this section, we present two case studies from the sentence production literature that we think successfully meet the criteria for using diverse languages to usefully broaden the empirical domain. The first is an example where the basic phenomenon has been studied before, but a cross-linguistic perspective provides a new point-of-view, namely ergative case. The second explores a linguistic construction, switch reference, which has previously received no attention in psycholinguistics.

4.1 Case study 1: Planning of sentences with ergative case marking

Case marking is a cross-linguistic strategy for signalling how nominal elements of a sentence relate to the verb and to each other. In the sample of the World Atlas of Language Structures, 52% of languages have some type of case system (Iggesen, 2013);

this, however, is also the source of enormous cross-linguistic variation. Ergative case alignment, for instance, is a common pattern of marking on argument nouns in which the agent of a transitive verb receives a case marker distinct from the case on objects or intransitive subjects (Bickel and Nichols, 2009). This differs from the case pattern found in all Indo-European languages in Europe, including English, which exhibit what is known as *nominative-accusative* alignment. In nominative-accusative alignment, the subjects of transitives and intransitives have the same case marking, and objects are the odd one out (being marked by accusative).

For nominative-accusative languages, it is well established that when describing a scene depicting a transitive event (i.e., one with an agent, a patient, and an action, such as a dog chasing a ball), speakers look to each element of the event in the order that they would be mentioned in a sentence (Griffin and Bock, 2000). This is a demonstration of incrementality in planning: producers do not wait until they have fully planned a sentence before they beginning their utterance, but rather plan the early parts of a sentence and begin speaking those before they have finished planning the later parts (see, e.g., Bock and Ferreira, 2014). Evidence from picture word interference studies dovetails with this finding, in showing that agents are planned independently at the beginning of the sentence, before the planning of verbs or patients (Momma and Ferreira, 2019; Momma et al., 2016). The preparation of a nominative initial noun phrase allows some flexibility in planning for the speaker (cf., e.g., Ferreira and Swets, 2002; Myachykov, Scheepers, Garrod, Thompson, and Fedorova, 2013) because it does not require the speaker to make any commitments about the existence of a patient or whether the verb is transitive or intransitive, providing additional time to plan the patient and verb while the agent is being pronounced.

Ergative case alignment presents a speaker with a different set of constraints, however, and therefore requires a different set of strategies. When planning a sentence in an ergative language, the choice of case marker on the initial noun (usually the agent)

depends on the transitivity of the verb (ergative if transitive, absolutive if intransitive). Given this dependency, speakers of ergative languages need to plan at least some information about the verb to select the appropriate case marker on the initial noun. They should, in other words, engage in more up-front relational planning (cf. Bock and Ferreira, 2014; Konopka, 2019) when formulating simple sentences, compared to speakers of nominative-accusative languages.

Egurtzegi, Blasi, Laka, Bornkessel-Schlesewsky, Meyer, Bickel, and Sauppe (2022a) provide evidence for such diverging planning strategies by comparing sentence planning in Basque (ergative agents) and Swiss German (nominative agents)in a picture description study. Using eye tracking, Egurtzegi, Blasi, Laka, Bornkessel-Schlesewsky, Meyer, Bickel, and Sauppe found that the peak proportion of looks to the agent was later in Basque than German, indicating that German speakers focused on the agent early, then moved on (to planning the verb and patient), while Basque speakers split their attention more between different aspects of the pictures during initial planning, which indicates increased relational encoding. Using EEG, Egurtzegi et al. found that differences in neural activity (in the theta, alpha, and beta frequency bands) also indicate that ergative case marking requires an earlier commitment to a sentence structure than nominative-accusative case marking.

This between-language difference in sentence planning has also been shown to apply within a single language that has a split case marking system and features both sentences with ergative-marked and with nominative-marked agents. Sauppe, Choudhary, Giroud, Blasi, Norcliffe, Bhattamishra, Gulati, Egurtzegi, Bornkessel-Schlesewsky, Meyer, and Bickel (2021) found that Hindi speakers looked back and forth more between the agent and other aspects of the pictures when planning ergative sentences. When planning sentences with nominative agents, Hindi speakers concentrated on the agent during the early phases of planning. Neural activity measured through EEG showed again that speakers committed to the structure under preparation earlier when planning sentences with ergative case marking because they need to commit to the transitivity of the verb

early in Hindi. This parallels the differences found in Egurtzegi et al.'s comparison of Basque and German sentence planning: Increased relational encoding presumably helps to establish which kind of verb should be used, which, in turn, determines whether the initial noun should carry nominative or ergative case. In sum, these studies show that what appears to be a relatively small morphosyntactic difference has substantial effects on processing.

4.2 Case study 2: Planning of sentences with switch reference marking

Our second case study focusses on switch reference, a linguistic phenomenon which has not, until recently, been studied psycholinguistically, but which provides a useful testing ground for refining theories of planning scope during sentence production. Switch reference is a system of chaining clauses together into a single sentence in which each (nonfinal) clause in the chain includes a morpheme (typically on the verb) that indicates whether the next clause has the same subject as the current clause or a different one (see van Gijn, 2016). The sentence in (1) provides an example of switch reference in Nungon, a Finisterre-Huon language spoken in Papua New Guinea.

(1) [Nan-na om-un-a], [ongo-ng-a], [Imom ir-a], [moro to-ng-a],
[father-1SG.POSS die-DS.3SG-MV] [go-DEP-MV] [Imom be-MV] [big do-DEP-MV]
[e-ng-a], [ngo-ndo ir-a], [amna to-ng hi-go-t].
[come-DEP-MV] [here-LDEM.NEAR be-MV] [man SG.O.take-DEP put-RP-1SG]
My father having died, (I) went on, stayed in Imom, became big, came (here, in Yawan), staying here, I took a man (as husband).

(adapted from Sarvasy, 2014, : 343)

The sentence in (1) has eight verbs, one of which has a third person subject (*om*, 'die') and the rest of which all have the first person speaker as the subject of their clause. Rather than indicating the identity of the subject in each individual clause, which would be repetitive for the seven verbs which share a subject, Nungon's particular instantiation of switch reference allows speakers to economize by simply dropping the subject verb

agreement when the subsequent verb will have the same subject (argument-drop for the subject is widely used as well, as it is in this example). This means that the only two verbs that need overt agreement are the last one, which marks the first person agreement for all seven first person clauses, and the very first clause which has *nan* 'father' as the subject. The agreement for this latter subject cannot be recovered anywhere else and therefore cannot be skipped. Consequently, the third person agreement morpheme *-un-* in the first clause doubles as an indicator that the next clause will have a new subject, while the lack of subject agreement on the second through seventh verbs functions as an indicator that in each the next verb has an identical subject.

Switch reference presents a challenge for current models of sentence production. Based on the languages studied so far, current theories take the clause as a major unit of production (Smith and Wheeldon, 1999), where the notion of a unit in this case implies that at some level the verb and its arguments are planned together. It is not immediately clear whether or how this unit size would apply to, say, the second verb in (1), *ongo* 'go' which is four clauses away from any realization of its argument.

In a visual world eye-tracking experiment, Sarvasy, Morgan, Yu, Ferreira, and Momma (2022) found that Nungon speakers shifted their attention in accordance with whether the current verb was marked to share the same subject or have a different subject from the subsequent clause. Moreover, speakers did so well before the onset of the morpheme itself, and therefore prior to the start of the subsequent clause (which could have been an alternative inflection point for gaze allocations). This maintains the tight link between gaze allocation and sentence planning in current models of psycholinguistics (Norcliffe and Konopka, 2015), but substantially expands the size of the planning window commonly assumed for sentence production. The evidence from Nungon thus provides a perspective on sentence planning that could not have been obtained from the languages usually represented in the psycholinguistics literature.

Taken together, the studies reviewed in this section show that crosslinguistic research

"can guide the development of theories by revealing the full extent of the human ability" to produce (and also comprehend) language (Jaeger and Norcliffe, 2009, p. 866).

5 RESEARCH METHODS AND RECOMMENDATIONS FOR PRACTICE

Cross-linguistic studies will, in many cases, need to be carried out outside of a university laboratory, in the places where the languages of interest are spoken or signed. This means that some experimental methods may not be used or that the available methods need to be adapted to the circumstances. Wagers and Chung (2019) elaborate on many aspects of designing and carrying out experiments "in the field". Speed, Wnuk, and Majid (2017) provide an overview of the best practices and pitfalls around cross-cultural experimental linguistic research and Whalen and McDonough (2015) also discuss aspects of field-based experiments, with a focus on phonetic data. Here, we discuss some additional requirements of cross-linguistic experimental research.

What is likely holding back many attempts of experimental psycholinguistic research outside of the laboratory is not any perceived lack of scientific interest, but rather practical concerns about data collection in sites away from the researchers' universities. Each language and the context and location of its speaker community poses its own challenges, including the logistics, the socio-cultural aspects of conducting experimental research, and the endangerment and documentation status of the languages (Seifart, Evans, Hammarström, and Levinson, 2018). Therefore, in-depth knowledge of the target language and the culture of its speakers is essential to successfully conduct experimental research. This means that intensive fieldwork and long-term engagement with the community is usually required, especially for lesser-described languages with small speaker communities. Collaboration with descriptive or field linguists working on the target language is often essential because the available linguistic descriptions may not

cover the phenomena of interest in sufficient detail or because there are interactions between grammatical domains which might otherwise be easy to overlook. At the same time, field linguists can profit from such a collaboration because it may provide them with the opportunity to more closely examine aspects of the target language that they have not yet studied (e.g., because a certain phenomenon did not yet occur in their materials). Næss and Sauppe (under review) provide a case report on a collaboration between descriptive linguists and psycholinguists about carrying out a sentence comprehension experiment on Äiwoo, spoken in the Solomon Islands. One question that arose during the preparation of this study was whether there are word order constraints on sentence-initial prepositional phrases, something that had not been documented in detail before. Collaborations between field linguists and psycholinguists must be "placed on an equal footing, ensuring that the standards of both disciplines are maintained" (Hellwig, 2019, p. 9), so that ideally both the experimental work and the description and documentation of the language are fostered. There are a number of comprehensive introductory and overview texts that provide further information on the practical, conceptual, and ethical issues surrounding linguistic fieldwork, such as Chelliah and de Reuse (2011), Thieberger (2012), and Bowern (2008).

Since only a small fraction of the world's languages have been studied experimentally, there are also many well-described ("large") languages that have not been comprehensively investigated by psycholinguistics yet. Surveying the grammatical and usage characteristics of these languages is facilitated by the availability of grammatical descriptions, linguistic research publications, and corpora of written and spoken language use and learning (Gries, 2012; Stoll and Schikowski, 2020). Among these larger languages are, for example, Indonesian and Swahili, but also Hindi (e.g., Choudhary, Schlesewsky, Roehm, and Bornkessel-Schlesewsky, 2009; Husain, Vasishth, and Srinivasan, 2015; Rubio-Fernandez and Jara-Ettinger, 2020), Tagalog (e.g., Pizarro-Guevara and Wagers, 2020; Sauppe, 2017a), or Arabic (e.g., Flecken, von Stutterheim, and Carroll, 2014; Matar, Pylkkänen, and

Marantz, 2019).

5.1 Stimulus materials

The creation of stimulus materials is also tied to knowledge of a language's grammar and use, as well as the cultural context of the speech community.

Studies on the processing of single words, for example, require in-depth knowledge about the structure of the lexicon. Studies on sentence comprehension require stimulus sentences with the intended target structures that must also be grammatical and felicitous in all other pragmatic and cultural aspects. For example, sentences like the classical English "lawyer" sentences in the study of relative clause processing (such as "The banker that irritated the lawyer/the lawyer irritated played tennis every Saturday"; Traxler, Morris, and Seely, 2002), while being grammatical, felicitous, and acceptable, can describe scenarios that are quite abstract or somewhat unusual. When preparing stimulus sentences for cross-linguistic studies, one needs to keep in mind that scenarios that are too abstract or unusual may cause difficulties in participants' understanding of the research, especially when it is hard for them to make sense of the task (cf. below).

Language processing studies that use pictures to track eye movements or elicit utterances or words (e.g., Mulak, Sarvasy, Tuninetti, and Escudero, 2021; Norcliffe, Konopka, Brown, and Levinson, 2015b; Nordlinger, Garrido Rodriguez, and Kidd, 2022; Sauppe et al., 2021) need pictures that are culturally appropriate and show concepts that are known to the participants. For example, someone who has never seen an elephant or jaguar may not recognise these animals and may have difficulty naming them. Participants also may not have experience with anthropomorphic animals, as they are commonly portrayed in cartoons, because as children they did not have access to television or other media, in contrast to the commonly studied Western university students. Garrido Rodriguez, Norcliffe, Huettig, Brown, and Levinson (in press) made photographs of objects that are in daily use or well known in a speaker community and used them in a comprehension

study to ensure that visual stimuli were easily recognised by participants. Such object photographs are also potentially useful for word production studies. To elicit sentences, depictions of (dynamic) actions are often used. Drawings are especially suitable for this purpose because they allow showing a large variety of scenarios and objects, which may be hard to photograph, for example, when featuring wild life (such as a lion hunting a zebra). However, a large number of actions involving two humans (e.g., pushing, kicking, feeding, or kissing) or a human and an inanimate object (e.g., peeling fruit, cutting paper, or watering a plant) can also be staged and photographed to create realistic and easily recognisable scenarios for eliciting descriptions in sentence form (for examples see Isasi-Isasmendi, Andrews, Flecken, Laka, Daum, Meyer, Bickel, and Sauppe, under revision; Sakarias and Flecken, 2019).

Experimental data that are contaminated by participants struggling to understand and contextualise culturally or linguistically incongruous materials may not be interpretable. However, this does not mean that stimulus sentences cannot describe novel or unusual scenarios or that pictures cannot show new or potentially unknown objects. Consulting with field linguists who have in-depth knowledge of the socio-cultural aspects of language use and with speakers of the target language are essential to find out what might or might not be experimentally felicitous. Testing the stimuli (pictures, sentences, and individual words, etc., depending on the experiment's aims) before commencing data collection is equally important to get a practical sense of what works well and what doesn't (especially since introspection and thinking about how certain stimuli will be perceived can turn out to be too pessimistic or too optimistic).

Since there are many factors that influence how language is processed, experimental studies often seek to control for at least some of those factors. At the level of individual words, for example, the lexical frequency and the phonological neighborhood density can affect recognition and production (Cutler, 2012; Vitevitch and Luce, 2016). When selecting the words to be used in the stimuli, one would want to take these factors

into account. The emergence and availability of large online corpora already allows the compilation of lexical frequency, but also constructional frequencies, for many languages, including lesser described ones: The Universal Dependencies database contains syntactic treebanks for more than 100 languages (de Marneffe, Manning, Nivre, and Zeman, 2021) and the Leipzig Corpora Collection hosts corpus-based monolingual dictionaries and downloadable corpora crawled from the internet for nearly 300 languages (`https://corpora.uni-leipzig.de/en`).

However, for languages with fewer resources, no internet-based corpora have been built. In some cases, it is possible to generate the necessary information, for example, about the frequency of constructions. To validate the observation that the most frequent word order in Äiwoo places the patient (or object) sentence-initially (Næss, 2015, 2021), Sauppe, Næss, Roversi, Meyer, Bornkessel-Schlesewsky, and Bickel (under revision) manually annotated the available corpus of glossed texts in the language (Næss, 2017) with syntactic information. These annotations allowed an estimate of how likely a patient-initial position was for each noun and utterance in the corpus. Although such *ad hoc* resources do not provide such broad coverage as corpora and databases based on hundreds of thousands or even millions of sentences, they nevertheless may provide the best available information on a target language and thus may crucially contribute to making stimulus creation (and statistical analyses) more comprehensive. Sometimes it is not possible or feasible to compile such information (e.g., when there are no text collections, or larger dictionaries don't exist). For frequency estimates, especially of individual words, it may be possible to obtain subjective frequency ratings (Brysbaert and Cortese, 2011; Thompson and Desrochers, 2009).

Especially for studies on small and understudied languages with rare linguistic features, the knowledge gain associated with an experimental study will in most cases outweigh the risks of not being able to control some aspects of the stimuli. When designing an experiment and preparing stimuli, the possible confounding factors should be considered,

and measures should be taken to mitigate them (as far as possible). For example, stimuli could be stratified based on available linguistic information (e.g., including both words that are ostensibly highly frequent and words that are ostensibly less frequent) or based on extra-linguistic factors (e.g., including picture stimuli that show easily recognisable actions and pictures that show more complex actions or that require more interpretation, cf. van de Velde, Meyer, and Konopka, 2014).

5.2 Participants

When studying a "large" language with many speakers, recruiting and testing participants may not differ much from how it would be done in Western universities. Especially if experiments can be carried out at a university, the student population can be invited to participate through posters or presentations in classes. However, when target languages are spoken by small or rural communities, the number of participants available may be small, and recruitment strategies must be tailored to the local situation (Speed et al., 2017), with the help of collaborators from the speaker community or a team member with in-depth knowledge of the local and socio-cultural circumstances. One possibility of engaging the community and inviting participation is to communicate the objectives and procedure of the experiment in (informally held) information meetings (Næss and Sauppe, under review).

Another possibility is to involve a native speaker collaborator as a "multiplier" to disseminate information about the research in the community to encourage other speakers to participate (Næss and Sauppe, under review). This approach presupposes that experimental linguists and psycholinguists, if they come to the community from outside, are already able to communicate the purpose and procedure in a way that inspires potential "multipliers". However, in general, it is important that participant recruitment is organised in a non-coercive way. Ethical approval should be obtained from the ethics committee or institutional review board of the researcher's university and also, if possible, from a local authority

such as a university in the area where the target language is spoken or from the (regional) government (Rice, 2012; Whalen and McDonough, 2015). This is also the case when targeting the community of speakers of a small language that may exist among the students of a university, e.g., in the capital of a region or country.

In addition, the plan for participant recruitment for experimental studies in the field needs to take into account the right time for data collection (e.g., the seasonality, such as avoiding the harvest season for communities living primarily from agriculture or taking into account the monsoon and other weather phenomena, Speed et al., 2017). It is also advisable to schedule enough time for data collection and prepare alternative recruitment approaches ahead of time to be able to change gears in reaction to the circumstances. For example, Næss and Sauppe (under review) intended to conduct their EEG study in the group of islands where Äiwoo is spoken but had to adapt and recruit participants in the capital of Solomon Islands because air travel was unexpectedly not available as planned and getting to the islands was not possible.

In recent years, web-based experiments have experienced an upswing, driven by improvements in technical accessibility and precision (Bridges, Pitiot, MacAskill, and Peirce, 2020; Crump, McDonnell, and Gureckis, 2013; Sauter, Draschkow, and Mack, 2020; Stewart, Chandler, and Paolacci, 2017) and probably further facilitated by the COVID-19 pandemic that prevented many researchers from collecting data in the laboratory or in the field. Garcia, Roeser, and Kidd (2022) describe how online experiments can be used for psycholinguistic studies with understudied languages, exemplified with a priming study on Tagalog (Garcia and Kidd, 2020). Online studies thus enable the study of languages with a larger speaker population with sufficient access to the internet and smartphones or computers. This development will make it increasingly more viable to conduct cross-linguistic psycholinguistic studies.

5.3 Experimental tasks and measurements

Although much of the research in psycholinguistics and experimental linguistics is based on written materials, such as judging the acceptability of written sentences or paradigms measuring reading behaviour to study word and sentence processing (e.g., eye tracking while reading, self-paced reading, or rapid serial presentation in EEG studies), many languages are only spoken or signed. Access to economic and educational participation also varies widely between speaker communities, so that literacy (in general and in the target language) may be low. For studies outside a laboratory, auditory stimuli and paradigms that require spoken responses are therefore often more suitable. For participants who are used to smartphones, responses by button press on a response pad or game controller may be unproblematic. For participants who are not used to button presses, the experimenter could consider only using vocal responses (e.g., in response to comprehension questions) or training participants to tap on a picture displayed on a tablet.

In any case, experiments need to be tailored for use in the field. People in non-WEIRD communities are usually not used to being tested and are therefore often not as socialised as "compliant responders" (Speed et al., 2017) as students in Western universities that typically participate in psycholinguistic studies (Arnett, 2008). Thus, the experimental tasks need to be adapted or contextualised so that they are clear to understand. If they are too opaque, it can be difficult to convey the purpose of the experiment, given the likely differences in common ground on the (implicit) expectations for experiments and linguistic research in general (Speed et al., 2017). Speaker communities of target languages for experiments often already have come in contact with descriptive linguistic research (as this is the main way the grammatical features of the language become known to the research community). Consequently, ideas about what language research "is about" may be centred on language documentation or the compilation of dictionaries and other text collections. Explaining the purpose and procedures of psycholinguistic

and experimental linguistic research is therefore important to generate common ground.

In principle, all the measurement techniques commonly used in psycholinguistics can also be used in experimental studies in the field. Paradigms that measure reaction times or accuracy, for example, in response to answering comprehension questions about stimulus sentences, require only a laptop (and possibly a microphone). The self-paced listening paradigm (Waters and Caplan, 2004; Waters, Caplan, and Yampolsky, 2003), for example, can provide a measure of the online time course of sentence comprehension without the need for additional equipment. Self-paced listening has been used with different age groups (Fallon, Peelle, and Wingfield, 2006; Suzuki, 2013), suggesting that it can also be a versatile tool for field-based studies. Another behavioural method that can give insight into online comprehension of sentences is "touch (or finger) tracking" (a variant of mouse tracking, cf., e.g., Kieslich, Henninger, Wulff, Haslbeck, and Schulte-Mecklenbeck, 2019; Spivey and Dale, 2006), where participants move objects on the screen of a tablet. In the study by Wagers, Borja, and Chung (2018), Chamorro speakers listened to ambiguous relative clauses and moved a small "puck" to one of two pictures corresponding to either interpretation of the relative clause as they were listening to the sentence, depending on how they comprehended it. Wagers et al. analysed the latency of the first initiation of a movement as an indicator of the online sentence parsing process.

Eye tracking devices and EEG devices have become increasingly mobile, so that they can usually be brought to the field site. Sauppe et al. (under revision) transported the equipment for an EEG experiment in a large backpack, comprising two laptops, a mobile EEG device (Neuroelectrics Enobio 32) and the necessary accessories, sound speakers, keyboards, a button response box, as well as power banks and solar panels. Yasunaga, Yano, Yasugi, and Koizumi (2015) also conducted a field-based EEG study, investigating sentence comprehension in Kaqchikel, a Mayan language of Guatemala.

Other neuroimaging techniques, such as near-infrared spectroscopy (NIRS, measuring local changes in hemoglobin oxygenation in the cortex with light sensors), are also

becoming increasingly mobile (cf., e.g., Lloyd-Fox, Papademetriou, Darboe, Everdell, Wegmuller, Prentice, Moore, and Elwell, 2014; Pinti, Tachtsidis, Hamilton, Hirsch, Aichelburg, Gilbert, and Burgess, 2020), with the potential to become a valuable neurophysiological method for field-based studies on language processing (cf. Minagawa and Cristia, 2019, for an overview of the technique applied to language processing research). An example of the application of functional NIRS is Koizumi, Takeshima, Tachibana, Asaoka, Saito, Niikuni, and Gyoba (2020), who used it to study the differences in planning subject-initial and verb-initial sentences in Kaqchikel.

Improvements in techniques such as webcam-based eye tracking (Vos, Minor, and Ramchand, 2022; Yang and Krajbich, 2021) make it possible to measure visual attention to picture or video stimuli remotely in online studies (cf. Garcia et al., 2022). With this technique, the "visual world paradigm" (Huettig, Rommers, and Meyer, 2011) can be used to study word recognition or sentence comprehension in understudied languages (for examples of visual world eye tracking studies *in situ* see, e.g., Garrido Rodriguez et al., in press; Mishra, Singh, Pandey, and Huettig, 2012; Sauppe, 2016). Although allowing only the use of stimuli in which areas of interest are placed relatively far apart due to their lower resolution and precision, webcam-based eye tracking may also function as a low-cost alternative to conventional research-grade eye trackers (which still cost at least several thousand dollars). Such an experimental setup only requires a laptop computer with a webcam and an internet connection or an additional computer that functions as the local server, which could enable the use of eye tracking for research teams who could otherwise not afford to buy a "research-grade" eye tracker.

Before starting a study in the field, the equipment to be used should be extensively piloted. This should ensure that the experimenter is highly confident in operating the devices (such as an eye tracker or an EEG) and is able to adapt settings or the study procedure if necessary. This could, for example, involve testing an EEG setup in the tropical house of the botanical garden if researchers based in colder climates plan to

collect data on a language spoken closer to the equator (and if it is already known that no air conditioning will be available).

Hellwig (2019) argues for the use of stimulus materials such as picture stories or objects to elicit "semi-structured" responses because these are arguably more easily integrated with descriptive and documentary linguistic work. Cross-linguistic research on sentence production has followed a variant of this approach by eliciting picture descriptions while measuring visual attention with eye trackers (Koizumi et al., 2020; Norcliffe et al., 2015b; Nordlinger et al., 2022; Sarvasy et al., 2022; Sauppe et al., 2021; Sauppe, Norcliffe, Konopka, Van Valin, and Levinson, 2013). Participants' descriptions are usually elicited freely, without restricting what they can say, but still guided by the semantic content of the picture stimuli. To study comprehension, sentence-picture matching tasks also use visual stimuli. In these tasks, participants hear a sentence while seeing two different pictures (e.g., of a dog biting a man and of a man biting a dog), needing to decide which one matches the sentence (Wagers, Borja, and Chung, 2015; Wagers et al., 2018).

Finally, the design of an experimental task also influences the amount of data that can be collected. Speakers of understudied languages may not be used to the "test-taking" and the potentially repetitive character of psycholinguistic experiments. In addition to finding a task that works well for most participants, the researcher must take into account how much data can be collected from each participant, balancing the availability of participants, the number of stimuli, and the length of the experimental session. It may be difficult to achieve the statistical power that should be aimed at in psycholinguistics, in general (Brysbaert, 2019; Vasishth, Mertzen, Jäger, and Gelman, 2018). However, some strategies can be considered to mitigate the power problem, such as dividing data collection into multiple sessions to collect more trials for each participant without making individual sessions too exhausting (cf. e.g., Smith and Little, 2018, for small N designs). Bayesian statistics allows one to explicitly take into account prior expectations

and knowledge for making inferences about a study's results (Dienes and Mclatschie, 2018; Kruschke and Liddell, 2018; Nicenboim and Vasishth, 2016; van de Schoot, Depaoli, King, Kramer, Märtens, Tadesse, Vannucci, Gelman, Veen, Willemsen, and Yau, 2021) and is therefore possibly more suitable for analysing smaller datasets. Sequential analysis designs make it possible to economise on the number of participants because they allow assessing the support for the hypothesis after each participant so that more data can be added if the support is inconclusive and data collection can be stopped once the desired level of statistical support is reached (Elsey, 2021; Lakens, 2014; Mani, Schreiner, Brase, Köhler, Strassen, Postin, and Schultze, 2021; Schönbrodt, Wagenmakers, Zehetleitner, and Perugini, 2017). Especially in the context of field-based research, where it is often difficult and expensive to obtain data, these approaches may help to successfully conduct and publish studies.

6 FUTURE DIRECTIONS

Given the historical focus on a small set of languages (Blasi et al., 2022; Jaeger and Norcliffe, 2009; Kidd and Garcia, 2022), any effort to collect data from understudied languages has the potential to add more knowledge to the field. Eventually, this will allow psycholinguistics and the language sciences to attribute cross-linguistic evidence the role it deserves in furthering our understanding of the usage, processing, and evolution of human language (Majid and Levinson, 2010).

7 FURTHER READING

Bowern, Claire. 2008. Linguistic fieldwork: A practical guide. London: Palgrave Macmillan.

This book discusses the challenges for linguistic research in the field and strategies to over come them, many of which (including issues of participant recruitment, navigating community dynamics, keeping equipment in the field, etc.) are as relevant to psycholinguistic research as to traditional documentation linguistics.

Henrich, J., Heine, S. J., and Norenzayan, A. (2010). The weirdest people in the world? Behavioral and Brain Sciences, 33:61–135.

This article presents the case for widening the set of populations studied across psychology, and provides the first articulation of a framework for understanding why western undergraduates might be an inappropriate model for broader human behavior.

Slobin, Dan Isaac. 2014. The crosslinguistic study of language acquisition: Volumes 1-5. Psychology Press.

This multi-volume collection provides an extensive overview of cross-linguistic experimental work done in the area of first language acquisition and language development.

Stoll, S. (2015). Studying language acquisition in different linguistic and cultural settings. In Bonvillain, N., editor, The Routledge Handbook of Linguistic Anthropology, chapter 10, pages 140–158. Routledge, New York.

This chapter gives an overview of the crosslinguistic factors influencing language acquisition and outlines the main research methods to study language development in crosslinguistic contexts.

Wagers, Matthew, and Sandra Chung. 2019. Language processing experiments in the

field. Manuscript. URL: https://escholarship.org/uc/item/5p5552vk.

This manuscript focuses on practical recommendations for conducting language experiments in the field, based on the authors' experience of doing psycholinguistic research on the Austronesian language Chomorro.

Whalen, D. H. and McDonough, J. (2015). Taking the laboratory into the field. Annual Review of Linguistics, 1:395–415.

This article reviews methods for linguistic research outside of the laboratory, with a focus on data collection for understudied languages.

Yang, Xiaozhi, and Ian Krajbich. 2021. Webcam-based online eye-tracking for behavioral research. Judgement and Decision Making 16:1485–1505.

This article introduces the possibility of eyetracking methods with relatively light technical equipment, therefore making it portable and easy to bring to a field site.

8 RELATED TOPICS

While psycholinguists treat differences across languages as independent variables that may explain processing differences between groups of speakers, typologists are interested in cross-linguistic variation as the explanandum (cf., e.g., Bickel, 2015). What explains the existence of recurrent patterns across languages? Why are some types of structures or categories rarer than others?

Language scientists from various approaches have long invoked cognitive and communicative pressures as an explanation for typological patterns (Bever and Langendoen, 1971; Christiansen

and Chater, 2008; Croft, 2003; Culbertson and Kirby, 2016; Gibson, Futrell, Piantadosi, Dautriche, Mahowald, Bergen, and Levy, 2019; Haspelmath, 2008; Hawkins, 1988, 2004; Jaeger and Tily, 2011; MacWhinney and Bates, 1989b; Mollica, Bacon, Zaslavsky, Xu, Regier, and Kemp, 2021; Schmidtke-Bode, 2009; Zipf, 1949). On this view, languages are less likely to have certain properties because they are harder to learn or process, or less suitable for communication. Various methods have been utilized to develop and test hypotheses in this area, including computer simulations (e.g., Christiansen and Kirby, 2003; Kirby, 1999; Steels, 2011), quantitative corpus studies (see Schnell and Schiborr, 2022, for a recent review), computational phylogenetic methods (e.g., Bickel, Witzlack-Makarevich, Choudhary, Schlesewsky, and Bornkessel-Schlesewsky, 2015; Dunn, Greenhill, Levinson, and Gray, 2011), and field-base structured data elicitation methods (e.g., Majid, Bowerman, van Staden, and Boster, 2007). In recent years, exciting progress has also been made using Artificial Language Learning paradigms (ALL), in which experimental participants are exposed to miniature languages that are designed to allow controlled comparison of properties of interest (for reviews, see, e.g., Fedzechkina, Newport, and Jaeger, 2016; Levshina, 2019). Using ALL paradigms, researchers can observe which kinds of structures are more accurately learnt (e.g., Culbertson, Gagliardi, and Smith, 2017) or how probabilistic distributions of alternative structures are regularised (e.g., Culbertson, Smolensky, and Legendre, 2012; Fedzechkina, Jaeger, and Newport, 2012) and from this draw links to typological patterns. In a related paradigm, manual gestures or sequences of gestures are elicited and similarities among the gestures are measured in order to infer cognitive biases in the absence of linguistic input (e.g., Culbertson, Schouwstra, and Kirby, 2020; Futrell, Hickey, Lee, Lim, Luchkina, and Gibson, 2015; Goldin-Meadow, Chee So, Özyürek, and Mylander, 2008; Schouwstra and de Swart, 2014).

References

Anand, Pranav, Sandra Chung, and Matthew Wagers. 2010. Widening the net: Challenges for gathering linguistic data in the digital age. *Response to NSF SBE* .

Aoshima, Sachiko, Colin Phillips, and Amy Weinberg. 2004. Processing filler-gap dependencies in a head-final language. *Journal of memory and language* 51:23–54.

Arnett, Jeffrey J. 2008. The neglected 95%: Why American psychology needs to become less American. *American Psychologist* 63:602–614.

Barthel, Mathias, and Sebastian Sauppe. 2019. Speech planning at turn transitions in dialog is associated with increased processing load. *Cognitive Science* 43.

Bates, Elizabeth, Antonella Devescovi, and Beverly Wulfeck. 2001. Psycholinguistics: A Cross-Language Perspective. *Annual Review of Psychology* 52:369–396.

Bever, Thomas G., and D. T. Langendoen. 1971. A dynamic model of the evolution of language. *Linguistic Inquiry* 2:433–463.

Bickel, Balthasar. 2015. Distributional typology: Statistical inquiries into the dynamics of linguistic diversity. In *The Oxford handbook of linguistic analysis*, ed. Bernd Heine and Heiko Narrog, chapter 37, 901–924. Oxford: Oxford University Press, 2nd edition.

Bickel, Balthasar, and Johanna Nichols. 2009. Case marking and alignment. In *The Oxford handbook of case*, ed. Andrej L. Malchukov and Andrew Spencer, chapter 20, 304–321. Oxford: Oxford University Press.

Bickel, Balthasar, Alena Witzlack-Makarevich, Kamal Kumar Choudhary, Matthias Schlesewsky, and Ina Bornkessel-Schlesewsky. 2015. The neurophysiology of language processing shapes the evolution of grammar: Evidence from case marking. *PLoS ONE* 10:e0132819.

Blasi, Damián E., Joseph Henrich, Evangelia Adamou, David Kemmerer, and Asifa Majid. 2022. Over-reliance on English hinders cognitive science. *Trends in Cognitive Sciences* 26:1153–1170.

Bock, Kathryn, and Victor S. Ferreira. 2014. Syntactically speaking. In *The Oxford handbook of language production*, ed. Matthew Goldrick, Victor S. Ferreira, and Michele Miozzo, chapter 2, 21–46. Oxford: Oxford University Press.

Bowern, Claire. 2008. *Linguistic fieldwork: A practical guide*. London: Palgrave Macmillan.

Bridges, David, Alain Pitiot, Michael MacAskill, and Jonathan Peirce. 2020. The timing mega-study: comparing a range of experiment generators, both lab-based and online. *PeerJ* 8:e9414.

Brysbaert, Marc. 2019. How many participants do we have to include in properly powered experiments? a tutorial of power analysis with reference tables. *Journal of Cognition* 2:16.

Brysbaert, Marc, and Michel J. Cortese. 2011. Do the effects of subjective frequency and age of acquisition survive better word frequency norms? *The Quarterly Journal of Experimental Psychology* 64:545–559.

Carminati, Maria Nella. 2002. *The processing of italian subject pronouns*. University of Massachusetts Amherst.

Chelliah, Shobhana L., and Willem J. de Reuse. 2011. *Handbook of descriptive linguistic fieldwork*. Heidelberg / London / New York: Springer.

Choudhary, Kamal Kumar, Matthias Schlesewsky, Dietmar Roehm, and Ina Bornkessel-Schlesewsky. 2009. The N400 as a correlate of interpretively relevant linguistic rules: Evidence from Hindi. *Neuropsychologia* 47:3012–3022.

Christiansen, Morten H., and Nick Chater. 2008. Language as shaped by the brain. *Behavioral and Brain Sciences* 31:489–558.

Christiansen, Morten H., and Simon Kirby. 2003. Language evolution: consensus and controversies. *Trends in Cognitive Sciences* 7:300–307.

Chung, Sandra, Manuel F Borja, and Matthew Wagers. 2012. Bridging methodologies: experimental syntax in the pacific. In *Linguistic Society of America 86th Annual Meeting*.

Clahsen, Harald. 2016. Contributions of linguistic typology to psycholinguistics. *Linguistic Typology* 20:599–614.

Croft, William. 2003. *Typology and Universals*. Cambridge: Cambridge University Press, 2nd edition.

Crump, Matthew J. C., John V. McDonnell, and Todd M. Gureckis. 2013. Evaluating Amazon's Mechanical Turk as a Tool for Experimental Behavioral Research. *PLoS ONE* 8:e57410.

Culbertson, Jennifer, Annie Gagliardi, and Kenny Smith. 2017. Competition between phonological and semantic cues in noun class learning. *Journal of Memory and Language* 92:343–358.

Culbertson, Jennifer, and Simon Kirby. 2016. Simplicity and specificity in language: Domain-general biases have domain-specific effects. *Frontiers in Psychology* 6:1964.

Culbertson, Jennifer, Marieke Schouwstra, and Simon Kirby. 2020. From the world to word order: Deriving biases in noun phrase order from statistical properties of the world. *Language* 96:696–717.

Culbertson, Jennifer, Paul Smolensky, and Géraldine Legendre. 2012. Learning biases predict a word order universal. *Cognition* 122:306–329.

Cutler, Anne. 1985. Cross-language psycholinguistics. *Linguistics* 23.

Cutler, Anne. 2012. *Native listening: Language experience and the recognition of spoken words*. Cambridge, MA: MIT Press.

Dienes, Zoltan, and Neil Mclatschie. 2018. Four reasons to prefer Bayesian analyses over significance testing. *Psychonomic Bulletin & Review* 25:207–218.

Dryer, Matthew S. 2013a. Expression of pronominal subjects. In *The world atlas of language structures online*, ed. Matthew S. Dryer and Martin Haspelmath. Leipzig: Max Planck Institute for Evolutionary Anthropology. URL `https://wals.info/chapter/101`.

Dryer, Matthew S. 2013b. Order of subject, object and verb. In *The world atlas of language structures online*, ed. Matthew S. Dryer and Martin Haspelmath. Leipzig: Max Planck Institute for Evolutionary Anthropology. URL `https://wals.info/chapter/81`.

Dunn, Michael, Simon J. Greenhill, Stephen C. Levinson, and Russell D. Gray. 2011. Evolved structure of language shows lineage-specific trends in word-order universals. *Nature* 473:79–82.

Egurtzegi, Aitor, Damián E. Blasi, Itziar Laka, Ina Bornkessel-Schlesewsky, Martin Meyer, Balthasar Bickel, and Sebastian Sauppe. 2022a. Cross-linguistic differences in case marking shape neural power dynamics and gaze behavior during sentence planning. *Brain and Language* 230:105127.

Egurtzegi, Aitor M., Damián E. Blasi, Itziar Laka, Ina Bornkessel-Schlesewsky, Martin Meyer, Balthasar Bickel, and Sebastian Sauppe. 2022b. Cross-linguistic differences in case marking shape neural power dynamics and gaze behavior during sentence planning. *Brain and Language* 230:105127.

Elsey, James. 2021. Powerful sequential designs using Bayesian estimation: A power analysis tutorial using brms, the tidyverse, and furrr. PsyArXiv.

Evans, Nicholas, and Stephen C. Levinson. 2009. The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences* 32:429–492.

Fallon, Marianne, Jonathan E. Peelle, and Arthur Wingfield. 2006. Spoken sentence processing in young and older adults modulated by task demands: Evidence from self-paced listening. *Journal of Gerontology B: Psychological Sciences* 61:P10–P17.

Fedzechkina, Maryia, T. Florian Jaeger, and Elissa L. Newport. 2012. Language learners restructure their input to facilitate efficient communication. *Proceedings of the National Academy of Sciences of the United States of America* 109:17897–17902.

Fedzechkina, Maryia, Elissa L. Newport, and T. Florian Jaeger. 2016. Miniature artificial language learning as a complement to typological data. In *The usage-based study of language learning and multilingualism*, ed. Lourdes Ortega, Andrea E. Tyler, Hae In Park, and Mariko Uno, chapter 12, 211–232. Georgetown University Press.

Ferreira, Fernanda, and Benjamin Swets. 2002. How incremental is language production? Evidence from the production of utterances requiring the computation of arithmetic sums. *Journal of Memory and Language* 46:57–84.

Ferreira, Victor S, and Hiromi Yoshita. 2003. Given-new ordering effects on the production of scrambled sentences in Japanese. *Journal of psycholinguistic research* 32:669–692.

Flecken, Monique, Christiane von Stutterheim, and Mary Carroll. 2014. Grammatical aspect influences motion event perception: findings from a cross-linguistic non-verbal recognition task. *Language and Cognition* 6:45–78.

Futrell, Richard, Tina Hickey, Aldrin Lee, Eunice Lim, Elena Luchkina, and Edward Gibson. 2015. Cross-linguistic gestures reflect typological universals: A subject-initial, verb-final bias in speakers of diverse language. *Cognition* 136:215–21.

Garcia, Rowena, and Evan Kidd. 2020. The acquisition of the tagalog symmetrical voice system: Evidence from structural priming. *Language Learning and Development* 16:399–425.

Garcia, Rowena, Jens Roeser, and Evan Kidd. 2022. Online data collection to address language sampling bias: lessons from the COVID-19 pandemic. *Linguistics Vanguard* .

Garrido Rodriguez, Gabriela, Elisabeth Norcliffe, Falk Huettig, Penelope Brown, and Stephen C. Levinson. in press. Anticipatory processing in a verb-initial mayan language: Eye-tracking evidence during sentence comprehension in Tseltal. *Cognitive Science* .

Gibson, Edward, Richard Futrell, Steven T. Piantadosi, Isabelle Dautriche, Kyle Mahowald, Leon Bergen, and Roger P. Levy. 2019. How efficiency shapes human language. *Trends in Cognitive Sciences* 23:389–407.

van Gijn, Erik. 2016. Switch reference: An overview. In *Switch reference 2.0*, ed. Erik van Gijn and Jeremy Hammond, 1–54. Amsterdam / Philadelphia: John Benjamins.

Goldin-Meadow, Susan, Wing Chee So, Aslı Özyürek, and Carolyn Mylander. 2008. The natural order of events: How speakers of different languages represent events nonverbally. *Proceedings of the National Academy of Sciences of the United States of America* 105:9163–9168.

Gries, Stefan Th. 2012. Corpus linguistics, theoretical linguistics, and cognitive/psycholinguistics: towards more and more fruitful exchanges. In *Corpus linguistics and variation in English: Theory and description*, ed. Joybrato Mukherjee and Magnus Huber, 41–63. Rodopi.

Griffin, Zenzi M, and Kathryn Bock. 2000. What the eyes say about speaking. *Psychological science* 11:274–279.

Haspelmath, Martin. 2008. Frequency vs. iconicity in explaining grammatical asymmetries. *Cognitive Linguistics* 19:1–33.

Hawkins, John A., ed. 1988. *Explaining language universals*. Oxford: Blackwell.

Hawkins, John A. 2004. *Efficiency and Complexity in Grammars*. Oxford: Oxford University Press.

Hellwig, Birgit. 2019. Linguistic diversity, language documentation and psycholinguistics: The role of stimuli. In *Methodological tools for linguistic description and typology*, ed. Aimée Lahaussois and Marine Vuillermet, number 16 in Language Documentation & Conservation Special Publication, chapter 2, 5–30. Honolulu: University of Hawai'i Press.

Henrich, Joseph, Steven J Heine, and Ara Norenzayan. 2010. The weirdest people in the world? *Behavioral and brain sciences* 33:61–83.

Huettig, Falk, Joost Rommers, and Antje S. Meyer. 2011. Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica* 137:151–171.

Husain, Samar, Shravan Vasishth, and Narayanan Srinivasan. 2015. Integration and prediction difficulty in Hindi sentence comprehension: Evidence from an eye-tracking corpus. *Journal of Eye Movement Research* 8:1–12.

Hwang, Heeju, and Elsi Kaiser. 2014. The role of the verb in grammatical function assignment in English and Korean. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 40:1363.

Iggesen, Oliver A. 2013. Number of cases. In *The world atlas of language structures online*, ed. Matthew S. Dryer and Martin Haspelmath. Leipzig: Max Planck Institute for Evolutionary Anthropology. URL `https://wals.info/chapter/49`.

Inoue, Atsu, and Janet Dean Fodor. 1995. Information-paced parsing of Japanese. In *Japanese sentence processing*, ed. Reiko Mazuka and Noriko Nagai, chapter 2. New York: Psychology Press.

Isasi-Isasmendi, Arrate, Caroline Andrews, Monique Flecken, Itziar Laka, Moritz M. Daum, Martin Meyer, Balthasar Bickel, and Sebastian Sauppe. under revision. The agent preference in visual event apprehension: probing the balance between cross-linguistic persistence and flexibility .

Jaeger, T. Florian, and Elisabeth J. Norcliffe. 2009. The cross-linguistic study of sentence production. *Language and Linguistics Compass* 3:866–887.

Jaeger, T. Florian, and Harry J. Tily. 2011. On language 'utility': processing complexity and communicative efficiency. *Wiley Interdisciplinary Reviews: Cognitive Science* 2:323–335.

Kamide, Yuki, Gerry TM Altmann, and Sarah L Haywood. 2003. The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and language* 49:133–156.

Kidd, Evan, and Rowena Garcia. 2022. How diverse is child language acquisition research? *First Language* 42:703–735.

Kieslich, Pascal J., Felix Henninger, Dirk U. Wulff, Jonas M. B. Haslbeck, and Michael Schulte-Mecklenbeck. 2019. Mouse-tracking: A practical guide to implementation and analysis. In *A handbook of process tracing methods*, ed. Michael Schulte-Mecklenbeck, Anton Kuehberger, and Joseph G. Johnson, chapter 8, 111–130. New York: Routledge, 2nd edition.

Kirby, Simon. 1999. *Function, selection, and innateness: The emergence of language universals*. Oxford / Cambridge, MA: Oxford University Press.

Koizumi, Masatoshi, Yasuhiro Takeshima, Ryo Tachibana, Riku Asaoka, Godai Saito, Keiyu Niikuni, and Jiro Gyoba. 2020. Cognitive loads and time courses related to word order preference in Kaqchikel sentence production: an NIRS and eye-tracking study. *Language, Cognition and Neuroscience* 35:137–150.

Konieczny, Lars, Barbara Hemforth, Christoph Scheepers, and Gerhard Strube. 1997. The role of lexical heads in parsing: Evidence from German. *Language and cognitive processes* 12:307–348.

Konopka, Agnieszka E. 2019. Encoding actions and verbs: Tracking the time-course of relational encoding during message and sentence formulation. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 45:1486–1510.

Kruschke, John K., and Torrin M. Liddell. 2018. The Bayesian New Statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic Bulletin & Review* 25:178–206.

Lakens, Daniël. 2014. Performing high-powered studies efficiently with sequential analyses. *European Journal of Social Psychology* 44:701–710.

Levshina, Natalia. 2019. Linguistic Frankenstein, or How to test universal constraints without real languages. In *Explanation in typology: Diachronic sources, functional motivations and the nature of the evidence*, ed. Karsten Schmidtke-Bode, Natalia Levshina, Susanne Maria Michaelis, and Ilja Seržant, number 3 in Conceptual Foundations of Language Science, chapter 9, 203–221. Language Science Press.

Levy, Roger P, and Frank Keller. 2013. Expectation and locality effects in German verb-final structures. *Journal of memory and language* 68:199–222.

Lloyd-Fox, Sarah, M. Papademetriou, M. K. Darboe, N. L. Everdell, R. Wegmuller, A. M. Prentice, S. E. Moore, and C.E. Elwell. 2014. Functional near infrared spectroscopy (fNIRS) to assess cognitive function in infants in rural Africa. *Scientific Reports* 4:4740.

MacWhinney, Brian, and Elizabeth Bates. 1989a. *The crosslinguistic study of sentence processing*. Cambridge University Press.

MacWhinney, Brian, and Elizabeth Bates, ed. 1989b. *The crosslinguistic study of sentence processing*. Cambridge: Cambridge University Press.

Majid, Asifa, Melissa Bowerman, Miriam van Staden, and James S. Boster. 2007. The semantic categories of cutting and breaking events: A crosslinguistic perspective. *Cognitive Linguistics* 18.

Majid, Asifa, and Stephen C. Levinson. 2010. WEIRD languages have misled us, too. *Behavioral and Brain Sciences* 33:193.

Mani, Nivedita, Melanie S. Schreiner, Julia Brase, Katrin Köhler, Katrin Strassen, Danilo Postin, and Thomas Schultze. 2021. Sequential Bayes Factor designs in developmental research: Studies on early word learning. *Developmental Science* 24:e13097.

de Marneffe, Marie-Catherine, Christopher D. Manning, Joakim Nivre, and Daniel Zeman. 2021. Universal dependencies. *Computational Linguistics* 47:255–308.

Matar, Suhail, Liina Pylkkänen, and Alec Marantz. 2019. Left occipital and right frontal involvement in syntactic category prediction: MEG evidence from Standard Arabic. *Neuropsychologia* 135.

Meyer, Antje S. 1996. Lexical access in phrase and sentence production: Results from picture–word interference experiments. *Journal of memory and Language* 35:477–496.

Minagawa, Yasuyo, and Alejandrina Cristia. 2019. Shedding light on language function and its development with optical brain imaging. In *The Oxford handbook of neurolinguistics*, ed. Greig I. de Zubicaray and Niels O. Schiller, chapter 7, 154–185. Oxford: Oxford University Press.

Mishra, Ramesh K., Niharika Singh, Aparna Pandey, and Falk Huettig. 2012. Spoken language-mediated anticipatory eye-movements are modulated by reading ability – Evidence from Indian low and high literates. *Journal of Eye Movement Research* 5:1–10.

Mitsugi, Sanako. 2017. Incremental comprehension of Japanese passives: Evidence from the visual-world paradigm. *Applied Psycholinguistics* 38:953–983.

Mollica, Francis, Geoff Bacon, Noga Zaslavsky, Yang Xu, Terry Regier, and Charles Kemp. 2021. The forms and meanings of grammatical markers support efficient communication. *Proceedings of the National Academy of Sciences of the United States of America* 118:e2025993118.

Momma, Shota, and Victor S Ferreira. 2019. Beyond linear order: The role of argument structure in speaking. *Cognitive psychology* 114:101228.

Momma, Shota, L Robert Slevc, and Colin Phillips. 2016. The timing of verb selection in Japanese sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 42:813.

Mulak, Karen E., Hannah S. Sarvasy, Alba Tuninetti, and Paola Escudero. 2021. Word learning in the field: Adapting a laboratory-based task for testing in remote Papua New Guinea. *PLoS ONE* 16:e0257393.

Myachykov, Andriy, Christoph Scheepers, Simon Garrod, Dominic Thompson, and Olga Fedorova. 2013. Syntactic flexibility and competition in sentence production: The case of English and Russian. *The Quarterly Journal of Experimental Psychology* 66:1601–1619.

Næss, Åshild. 2015. The Äiwoo verb phrase: Syntactic ergativity without pivots. *Journal of Linguistics* 51:75–106.

Næss, Åshild. 2017. Documenting Äiwoo. Endangered Languages Archive. URL `http://hdl.handle.net/2196/00-0000-0000-000F-BF44-A`.

Næss, Åshild. 2021. Voice and valency morphology in Äiwoo. *Oceanic Linguistics* 60:160–198.

Næss, Åshild, and Sebastian Sauppe. under review. Bringing psycholinguistics to the field: experiences from Solomon Islands. *Language Documentation & Conservation* .

Nakano, Yoko, Claudia Felser, and Harald Clahsen. 2002. Antecedent priming at trace positions in Japanese long-distance scrambling. *Journal of Psycholinguistic Research* 31:531–571.

Nicenboim, Bruno, and Shravan Vasishth. 2016. Statistical methods for linguistic research: Foundational ideas — part II. *Language and Linguistics Compass* 10:591–613.

Nielsen, Mark, Daniel B. M. Haun, Joscha Kärtner, and Cristine H. Legare. 2017. The persistent sampling bias in developmental psychology: A call to action. *Journal of Experimental Child Psychology* 162:31–38.

Norcliffe, Elisabeth, Alice C. Harris, and T. Florian Jaeger. 2015a. Cross-linguistic psycholinguistics and its critical role in theory development: Early beginnings and recent advances. *Language, Cognition and Neuroscience* 30:1009–1032.

Norcliffe, Elisabeth, and Agnieszka E. Konopka. 2015. Vision and language in cross-linguistic research on sentence production. In *Attention and vision in language processing*, ed. Ramesh K. Mishra, Narayanan Srinivasan, and Falk Huettig, chapter 5, 77–96. New Delhi: Springer.

Norcliffe, Elisabeth, Agnieszka E. Konopka, Penelope Brown, and Stephen C. Levinson. 2015b. Word order affects the time course of sentence formulation in Tzeltal. *Language, Cognition and Neuroscience* 30:1187–1208.

Nordlinger, Rachel, Gabriela Garrido Rodriguez, and Evan Kidd. 2022. Sentence

planning and production in Murrinhpatha, an Australian 'free word order' language. *Language* 98:187–220.

Pinti, Paola, Ilias Tachtsidis, Antonia Hamilton, Joy Hirsch, Clarisse Aichelburg, Sam Gilbert, and Paul W. Burgess. 2020. The present and future use of functional near-infrared spectroscopy (fNIRS) for cognitive neuroscience. *Annals of the New York Academy of Sciences* 1464:5–29.

Pizarro-Guevara, Jed Sam, and Matthew Wagers. 2020. The predictive value of Tagalog voice morphology in filler-gap dependency formation. *Frontiers in Psychology* 11:517.

Rice, Keren. 2012. Ethical issues in linguistic fieldwork. In Thieberger (2012), chapter 18, 407–428.

Rubio-Fernandez, Paula, and Julian Jara-Ettinger. 2020. Incrementality and efficiency shape pragmatics across languages. *Proceedings of the National Academy of Sciences of the United States of America* 117:13399–13404.

Runner, Jeffrey T, and Alyssa Ibarra. 2016. Information structure effects on null and overt subject comprehension in Spanish. *Empirical perspectives on anaphora resolution* 87–112.

Saah, Kofi Korankye. 1995. Studies in Akan syntax, acquisition, and sentence processing. Doctoral Dissertation, University of Ottawa (Canada).

Sakarias, Maria, and Monique Flecken. 2019. Keeping the result in sight and mind: General cognitive principles and language-specific influences in the perception and memory of resultative events. *Cognitive Science* 43:e12708.

Saleem, Safiyyah. 2010. Argument optionality: A new library for the Grammar Matrix customization system. Doctoral Dissertation, University of Washington.

Sarvasy, Hannah S, Adam Milton Morgan, Jenny Yu, Victor S Ferreira, and Shota

Momma. 2022. Cross-clause planning in Nungon (Papua New Guinea): Eye-tracking evidence. *Memory & Cognition* 1–15.

Sarvasy, Hannah Sacha. 2014. A grammar of Nungon: A Papuan language of the Morobe province, Papua New Guinea. Doctoral Dissertation, James Cook University.

Sauppe, Sebastian. 2016. Verbal semantics drives early anticipatory eye movements during the comprehension of verb-initial sentences. *Frontiers in Psychology* 7.

Sauppe, Sebastian. 2017a. Symmetrical and asymmetrical voice systems and processing load: Pupillometric evidence from sentence production in Tagalog and German. *Language* 93:288–313.

Sauppe, Sebastian. 2017b. Word order and voice influence the timing of verb planning in German sentence production. *Frontiers in psychology* 8:1648.

Sauppe, Sebastian, Kamal Kumar Choudhary, Nathalie Giroud, Damián E. Blasi, Elisabeth Norcliffe, Shikha Bhattamishra, Mahima Gulati, Aitor Egurtzegi, Ina Bornkessel-Schlesewsky, Martin Meyer, and Balthasar Bickel. 2021. Neural signatures of syntactic variation in speech planning. *PLOS Biology* 19:e3001038.

Sauppe, Sebastian, Åshild Næss, Giovanni Roversi, Martin Meyer, Ina Bornkessel-Schlesewsky, and Balthasar Bickel. under revision. An agent-first preference in a patient-first language during sentence comprehension .

Sauppe, Sebastian, Elisabeth J. Norcliffe, Agnieszka E. Konopka, Robert D. Van Valin, and Stephen C. Levinson. 2013. Dependencies first: Eye tracking evidence from sentence production in Tagalog. In *Proceedings of the 35th Annual Conference of the Cognitive Science Society*, ed. Markus Knauff, Michael Pauen, Natalie Sebanz, and Ipke Wachsmuth, 1265–1270. Austin, TX: Cognitive Science Society.

Sauter, Martin, Dejan Draschkow, and Wolfgang Mack. 2020. Building, hosting and recruiting: A brief introduction to running behavioral experiments online. *Brain Sciences* 10:251.

Schmidtke-Bode, Karsten. 2009. *A Typology of Purpose Clauses*, volume 88 of *Typological Studies in Language*. Amsterdam / Philadelphia: John Benjamins.

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

Schönbrodt, Felix D., Eric-Jan Wagenmakers, Michael Zehetleitner, and Marco Perugini. 2017. Sequential hypothesis testing with bayes factors: Efficiently testing mean differences. *Psychological Methods* 22:322–339.

van de Schoot, Rens, Sarah Depaoli, Ruth King, Bianca Kramer, Kaspar Märtens, Mahlet G. Tadesse, Marina Vannucci, Andrew Gelman, Duco Veen, Joukje Willemsen, and Christopher Yau. 2021. Bayesian statistics and modelling. *Nature Reviews Methods Primers* 1:1.

Schouwstra, Marieke, and Henriëtte de Swart. 2014. The semantic origins of word order. *Cognition* 131:431–436.

Schriefers, Herbert, Encarna Teruel, and Raik-Michael Meinshausen. 1998. Producing simple sentences: Results from picture–word interference experiments. *Journal of Memory and Language* 39:609–632.

Seifart, Frank, Nicholas Evans, Harald Hammarström, and Stephen C. Levinson. 2018. Language documentation twenty-five years on. *Language* 94:e324–e345.

Sekerina, Irina A. 1997. *The syntax and processing of scrambling constructions in Russian*. City University of New York.

Sekerina, Irina A. 2003. Scrambling and processing: Dependencies, complexity, and constraints. *Word order and scrambling* 4:301–324.

Slobin, Dan Isaac. 2014. *The crosslinguistic study of language acquisition: Volumes 1-5*. Psychology Press.

Smith, Mark, and Linda Wheeldon. 1999. High level processing scope in spoken sentence production. *Cognition* 73:205–246.

Smith, Philip L., and Daniel R. Little. 2018. Small is beautiful: In defense of the small-*n* design. *Psychonomic Bulletin & Review* 25:2083–2101.

Speed, Laura, Ewelina Wnuk, and Asifa Majid. 2017. Studying psycholinguistics out of the lab. In *Research methods in psycholinguistics and the neurobiology of language: A practical guide*, ed. Annette M. B. de Groot and Peter Hagoort, chapter 10, 190–207. Hoboken, NJ: Wiley-Blackwell.

Spivey, Michael J., and Rick Dale. 2006. Continuous dynamics in real-time cognition. *Current Directions in Psychological Science* 15:207–211.

Steels, Luc. 2011. Modeling the cultural evolution of language. *Physics of Life Reviews* 8:339–356.

Stewart, Neil, Jesse Chandler, and Gabriele Paolacci. 2017. Crowdsourcing samples in cognitive science. *Trends in Cognitive Sciences* 21:736–748.

Stoll, Sabine. 2009. Crosslinguistic approaches to language acquisition. In *The cambridge handbook of child language*, ed. Edith L. Bavin, chapter 6, 89–104. Cambridge: Cambridge University Press.

Stoll, Sabine, and Robert Schikowski. 2020. Child-language corpora. In *A practical handbook of corpus linguistics*, ed. Magali Paquot and Stefan Th. Gries, chapter 14, 305–327. Cham: Springer.

Suzuki, Takaaki. 2013. Children's on-line processing of scrambling in Japanese. *Journal of Psycholinguistic Research* 42:119–137.

Tamaoka, Katsuo, Hiromu Sakai, Jun-ichiro Kawahara, Yayoi Miyaoka, Hyunjung Lim, and Masatoshi Koizumi. 2005. Priority information used for the processing of Japanese sentences: Thematic roles, case particles or grammatical functions? *Journal of Psycholinguistic Research* 34:281–332.

Thieberger, Nicholas, ed. 2012. *The Oxford handbook of linguistic fieldwork*. Oxford: Oxford University Press.

Thompson, Glenn L., and Alain Desrochers. 2009. Corroborating biased indicators: Global and local agreement among objective and subjective estimates of printed word frequency. *Behavior Research Methods* 41:452–471.

Traxler, Matthew J., Robin K. Morris, and Rachel E. Seely. 2002. Processing subject and object relative clauses: Evidence from eye movements. *Journal of Memory and Language* 47:69–90.

Vasishth, Shravan, Daniela Mertzen, Lena A. Jäger, and Andrew Gelman. 2018. The statistical significance filter leads to overoptimistic expectations of replicability. *Journal of Memory and Language* 103:151–175.

van de Velde, Maartje, Antje S. Meyer, and Agnieszka E. Konopka. 2014. Message formulation and structural assembly: Describing "easy" and "hard" events with preferred and dispreferred syntactic structures. *Journal of Memory and Language* 71:124–144.

Vitevitch, Michael S., and Paul A. Luce. 2016. Phonological neighborhood effects in spoken word perception and production. *Annual Review of Linguistics* 2:75–94.

Vos, Myrte, Serge Minor, and Gillian Catriona Ramchand. 2022. Comparing infrared and webcam eye tracking in the Visual World Paradigm. *Glossa Psycholinguistics* 1.

Wagers, Matthew, Manuel F. Borja, and Sandra Chung. 2015. The real-time comprehension of WH-dependencies in a WH-agreement language. *Language* 91:109–144.

Wagers, Matthew, and Sandra Chung. 2019. Language processing experiments in the field. Manuscript. URL `https://escholarship.org/uc/item/5p5552vk`.

Wagers, Matthew W., Manuel F. Borja, and Sandra Chung. 2018. Grammatical licensing and relative clause parsing in a flexible word-order language. *Cognition* 178:207–221.

Waters, Gloria S., and David Caplan. 2004. Verbal working memory and on-line syntactic processing: Evidence from self-paced listening. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology* 57:129–163.

Waters, Gloria S., David Caplan, and Sasha Yampolsky. 2003. On-line syntactic processing under concurrent memory load. *Psychonomic Bulletin & Review* 10:88–95.

Whalen, D. H., and Joyce McDonough. 2015. Taking the laboratory into the field. *Annual Review of Linguistics* 1:395–415.

Yamashita, Hiroko. 1997. The effects of word-order and case marking information on the processing of Japanese. *Journal of psycholinguistic research* 26:163–188.

Yang, Xiaozhi, and Ian Krajbich. 2021. Webcam-based online eye-tracking for behavioral research. *Judgement and Decision Making* 16:1485–1505.

Yasunaga, Daichi, Masataka Yano, Yoshiho Yasugi, and Masatoshi Koizumi. 2015. Is the subject-before-object preference universal? an event-related potential study in the Kaqchikel Mayan language. *Language, Cognition and Neuroscience* 30:1209–1229.

Zipf, George Kingsley. 1949. *Human behavior and the principle of least effort: an introduction to human ecology*. Cambridge, MA: Addison-Wesley Press.