John Benjamins Publishing Company

This is a contribution from SL 46:2
© 2022. John Benjamins Publishing Company

This electronic file may not be altered in any way. The author(s) of this article is/are permitted to use this PDF file to generate printed copies to be used by way of offprints, for their personal use only.

Permission is granted by the publishers to post this file on a closed server which is accessible only to members (students and faculty) of the author's/s' institute. It is not permitted to post this PDF on the internet, or to share it on sites such as Mendeley, ResearchGate, Academia.edu.

Please see our rights policy on *https://benjamins.com/content/customers/rights*
For any other use of this material prior written permission should be obtained from the publishers or through the Copyright Clearance Center (for USA: *www.copyright.com*).

Please contact *rights@benjamins.nl* or consult our website: *www.benjamins.com*

Predicting voice choice in symmetrical voice languages

All the things that do not work in Totoli

Sonja Riesberg[1,2,3] Maria Bardají i Farré[1] Kurt Malcher[1] and Nikolaus P. Himmelmann[1]
[1] Universität zu Köln | [2] CNRS–LACITO | [3] The Australian National University CoEDL

Western Austronesian symmetrical voice languages exhibit at least two basic transitive constructions. This paper investigates what factors influence speakers' choice of one voice over another in natural spoken discourse. It provides a thorough assessment of all factors that have been proposed to be relevant for voice choice in the literature on symmetrical voice systems. Using the Indonesian language Totoli as a case study, we show that unlike in languages with asymmetrical voice alternations, argument-related properties such as topicality, activation state, animacy, etc. do not play a major role in voice choice in symmetrical voice languages. We argue that for symmetrical voice alternations other factor groups are relevant than for asymmetrical voice alternations and that the clear structural differences between the two alternation types are mirrored in functional differences.

Keywords: western Austronesian, symmetrical voice, voice choice, discourse, Totoli

1. Introduction

Western Austronesian languages are famous for their typologically rare voice systems, which allow them to express all transitive eventualities with at least two syntactically equally transitive clause structures and which have therefore been called **symmetrical voice systems** (in the earlier literature, these are also known as **Philippine-type focus systems**). Consider the following two examples from

Additional material available from https://doi.org/10.1075/sl.20061.rie.additional
https://doi.org/10.1075/sl.20061.rie | Published online: 27 September 2021
Studies in Language 46:2 (2022), pp. 453–516. ISSN 0378-4177 | E-ISSN 1569-9978
© John Benjamins Publishing Company

Totoli, an Austronesian symmetrical voice language spoken in Sulawesi (Indonesia):[1]

(1) a. *I Ali nambalung deuk itu.*
i Ali noN-mbalung deuk itu
HON PN AV.RLS-throw.at dog DIST
'Ali threw (stones) at the dog.'

b. *Deuk itu ni-mbalung i Ali.*
dog DIST UV.RLS-throw.at HON PN
'Ali threw (stones) at the dog.'

From a cross-linguistic perspective, the data in (1) are remarkable, because they illustrate a voice alternation – an alternation that allows for the linking of different semantic roles to the subject relation – without at the same time changing the transitivity of the clause and demoting one of the arguments to an oblique function. In (1a), the NP in subject function, *i Ali*, is an actor and the verb is marked by actor voice morphology (in this example by the actor voice realis prefix *noN-*). If the subject is an undergoer, such as *deuk itu* in (1b), undergoer voice morphology has to be used (here the undergoer voice realis prefix *ni-*). In both actor voice and undergoer voice, the non-subject argument has core argument status, i.e. both clauses are syntactically fully transitive. See Riesberg (2014) for further discussion and details.

Due to their symmetry, western Austronesian voice systems pose challenges to language description and grammatical theory. For example, the concept of non-demoting voice alternations has proved to be challenging for both formal syntactic theories and typological-descriptive frameworks (see Chen & McDonnell 2019 for a recent review). Here, however, our main concern is with another obvious question: How do speakers choose one transitive construction over the other? In

1. Examples are represented according to the following conventions. Elicited data are signaled by the application of punctuation and capitalization both in the vernacular text and the translation. Spontaneous spoken data are given in intonation units (one unit per line) and do not apply sentence-final punctuation or sentence-initial capitalization. In examples from spoken texts, we also provide a label indicating the recording name and the number of the intonation unit.
The capital letter *N* in the prefix *noN-* represents a nasal that assimilates to, and in some cases substitutes for, the initial consonant of the stem. Before a vowel, it is realized as a velar nasal. See Himmelmann (2005: 118–120) and Bracks (2020) for further details. Angle brackets < > indicate false starts. Examples that involve morpho-phonologically complex processes are represented in four lines (vernacular, morpheme breaks, interlinear gloss, free translation); all other examples have only three lines (morpheme breaks are indicated directly in the vernacular line).
All data used for the present study are available in the Totoli documentation collections by Leto et al. (2005–2010) and Bracks et al. (2017–2020).

© 2022. John Benjamins Publishing Company
All rights reserved

asymmetrical voice alternations such as the active-passive alternation, one of the voices has a marked status in that it is only used in special circumstances and typically also much less frequently than the other, unmarked voice. Thus, the passive in English, if accompanied by an agent phrase, is typically used when the undergoer has a higher level of givenness than the actor (Biber et al. 1999: 941). In the case of symmetrical voice languages, the question of which factors determine voice choice is still wide open, as we will show in this paper.

This question has occupied typologists and Austronesianists since the late 1970s, after Schachter (1976, 1977) had made the western Austronesian voice system a major concern in the debate on grammatical relations. Especially the notions of **discourse topicality** (Givón 1983) and **discourse transitivity** (Hopper & Thompson 1980) have inspired a number of works that address these topics for Austronesian symmetrical voice languages (e.g. Cooreman (1983, 1988) for Chamorro, Cumming (1991) for Classical Malay, Hopper (1983) for Malay, McDonnell (2016) for Besemah, Pastika (1999) for Balinese, Quick (2005) for Pendau, Wouk (1996) for Jakarta Indonesian, and Wouk (1999) and Asikin-Garmager (2017) for Sasak). In these works, a number of factors has been proposed to be of relevance for voice choice in symmetrical voice languages. Our goal here is to provide a comprehensive survey of all the factors that have been suggested and to test their relevance on corpus data from Totoli. This fresh attempt to tackle the voice choice issue in symmetrical voice languages is motivated by the following observations:

- The literature to date is mostly based on narrative data, with a few exceptions such as Wouk (1999) and McDonnell (2016) who, however, only deal with conversational data. Some of the older literature makes use of written data or mixes spoken and written data. More often than not the sample size is quite small (sometimes less than 150 transitive clauses; e.g. Brainard 1994, Cumming 1991, Hopper 1983) and hardly suited for proper statistical inquiry. No work to date is based on a corpus that includes conversational as well monological speech and also provides different genres for each of these two basic types of linguistic interaction. Among others, Hemmings (2016) emphasizes the need to control for genre and shows that genre plays a major role for the choice among alternative word orders (Section 5.3 in Hemmings 2016). As detailed in Section 2, our Totoli corpus is designed to meet basic requirements for sample size and diversity of text types.
- In the literature to date, there has been no attempt to test a comprehensive set of factors for voice choice, most works dealing at most with two or three factors. The major exception is McDonnell's (2016) thorough investigation of Besemah, which not only covers the major factors discussed in

© 2022. John Benjamins Publishing Company
All rights reserved

the older literature but also includes a substantial number of new factors such as the presence of VALENCY-INCREASING MORPHOLOGY, COLLOSTRUCTION STRENGTH and SYNTACTIC PRIMING. However, McDonnell's study is based exclusively on everyday conversation and its results are difficult to compare with the earlier literature. Furthermore, INTERACTIVITY and TEXT TYPE are not investigated as possible factors.

- The earlier literature does not make use of state-of-the-art statistical techniques such as generalized linear mixed-effects models. McDonnell (2016) once again is an exception who, however, does not consider the effect size or relative importance of the factors he investigates.
- In terms of morphosyntactic marking (the marking of core arguments and the voice morphology on the verb) and the syntactic properties of the core arguments, Totoli is the most consistent symmetrical voice language documented to date (Riesberg 2014).[2] It thus promises to provide the clearest pointers to the factors involved in voice choice in these languages.

Given the shortcomings of the earlier literature just mentioned, but also given the fact that comparatively little work on this topic is available in the first place, it should not come as a surprise that the factors determining voice choice in symmetrical voice languages have remained unclear to date. What is also unclear is whether this state of affairs is primarily due to the different methodologies, conceptual frameworks and data bases that have been used in earlier work, or whether one or more important factors have been missed. Our investigation points to the latter option.

In this regard, it will be useful to remember that most factors that have been investigated so far are inspired by investigations of **asymmetrical** voice alternations, in particular the active-passive alternation in modern European languages. For these alternations it is clear that voice choice is closely linked to subject choice. In rather simplistic terms, there is a strong tendency for the more topical and/or more highly activated core argument to become the subject. The strong prevalence for the active voice then results from two factors. First, the actor tends to be more topical and more highly activated than the undergoer in many contexts. Second, in those instances where actor and undergoer are almost equal with regard to all relevant parameters (in addition to topicality and givenness, for example, humanness or distance to last mention), there is a default preference

2. Specifically, the non-subject arguments in both actor and undergoer voice share almost identical grammatical properties, while in other symmetrical voice languages, this is not the case to the same extent. For example, there are no definiteness or specificity constraints regarding the non-subject argument of an actor voice verb in Totoli. Riesberg (2014: 81–85) provides further details and examples.

© 2022. John Benjamins Publishing Company
All rights reserved

for actor subjects. See Svartvik (1966), Biber et al. (1999) and Zúñiga & Kittilä (2019: 89f) for details and exemplification.

Given that symmetrical voice systems also allow for alternative options in linking semantic roles to subject function, one could hypothesize that subject-related properties are of primary importance in determining voice choice in these systems as well. This would include the possibility that the relevant factors are essentially the same, but that their relative importance differs (e.g. humanness could be more important than topicality). Alternatively, factors that are not directly related to inherent or discourse-related properties of the core arguments could be more relevant than factors that are. While subject status changes in symmetrical voice alternations, core argument status does not (both actor and undergoer remain core arguments). This difference vis-a-vis asymmetrical voice systems may go hand in hand with major functional differences. Consequently, we can rephrase our main research question – how do speakers of symmetrical voice languages choose one transitive construction over the other? – in the following, more precise, manner:

1. To what extent is voice choice in symmetrical voice languages motivated by subject choice? To what extent does it rely on argument-related properties?
2. Is the major formal difference between symmetrical and asymmetrical voice alternations also reflected in a functional one? Does this include new factors not found to be relevant for asymmetrical alternations? Or are the differences confined to differing choices made with regard to factors also relevant in asymmetrical alternations?

Answering these questions requires us to thoroughly investigate all the factors that have been suggested as relevant in voice choice so far. Sections 3 to 6, the empirical core of this paper, are devoted to this at times somewhat tedious but unavoidable task. They provide both a brief summary of the literature and an evaluation of the extent to which a particular factor appears to play a role in the Totoli data. Given the state of the art, the factor group discussed in Section 3 is of particular relevance as it pertains to inherent and discourse-related properties of the core arguments such as animacy (inherent) and topicality (discourse related). The remaining sections explore factors that have played a lesser role in the literature, as they are not concerned with argument-related properties. The factors discussed in Section 4 pertain to the verb (e.g. presence of valency-increasing morphology and collostruction strength) or the clause (grounding). Section 5 concerns priming, and Section 6 deals with text and interactivity types.

The review of the literature in these sections will show that for many factors and factor groups the definitions used in earlier work vary to some degree, and sometimes are either not or not very easily operationalizable. Thus, while we try

© 2022. John Benjamins Publishing Company
All rights reserved

to be truly comprehensive in our assessment of previously discussed factors for voice choice, there are limits to this endeavour – in some instances it will not be possible to replicate exactly the procedures used in a previous study.

Throughout Sections 3–6, the statistical significance of each of the factors is assessed individually by means of generalized linear mixed-effects models. In Section 7, we compare the factors more directly with regard to the impact that they appear to have on voice choice. Specifically, we discuss the effect size and the relative importance of the factors investigated, estimated by means of Cramér's V coefficients (Cramér 1946) and a decision tree (Hothorn et al. 2006; Strobl et al. 2009).

Section 8 summarizes our results. The main result is a negative answer to the first group of questions above, given in (1). Subject choice does not appear to play a major role when choosing one or the other symmetrical voice. In fact, we find that the factors reviewed in Sections 3–6 do not predict voice distribution in the data satisfactorily. This points to the need to further investigate possible interactions among the factors to see whether a particular combination of them may allow for stronger predictions. But it also points to the possibility that factors not yet investigated may be of relevance. In this regard, we provide preliminary evidence for the hypothesis that episode boundaries may play a significant role in symmetrical voice choice.

To conclude this introduction, a brief note on grammatical relations in symmetrical voice languages is in order. Following Schachter (1976), there has been a vigorous debate as to whether the grammatical relation **subject** can be identified in symmetrical voice languages. The majority view in the specialist literature currently tends towards an affirmative answer to this question, following in particular the argument in Kroeger (1993). We basically agree with this view and use the term "subject" here as equivalent to what is termed "privileged syntactic argument" (PSA) in Van Valin (2005) and elsewhere. A PSA is defined as the syntactic element that controls coding properties such as agreement and is the pivotal element in complex constructions such as relativization, NP deletion, control, etc. In our usage, a "subject" is thus a language-specific and constructionally defined concept. Importantly, however, nothing in the current argument actually depends on this terminological choice. The data we discuss here once again show that the voice alternations in symmetrical voice languages clearly differ from the active-passive alternation well known from European languages. This holds not only for their structural properties – much discussed in the grammatical relations literature – but also for the discourse conditions determining the choice between actor and undergoer voice.

Before investigating these conditions, we start with a short introduction covering relevant aspects of Totoli grammar as well as details regarding the corpus assembled for this study and the set of annotations used in the statistical analyses.

© 2022. John Benjamins Publishing Company
All rights reserved

2. Language, corpus and methods

2.1 Totoli

Our study investigates voice choice in Totoli, a Western Malayo-Polynesian language spoken by up to 5000 speakers in the northern part of Central Sulawesi. See Himmelmann (2001, 2010) for further details on the social and linguistic setting.

As a symmetrical voice language, Totoli has two basic transitive constructions: actor voice and undergoer voice. Actor voice predicates have an actor argument as the subject, whereas in undergoer voice constructions the argument in subject function is an undergoer.[3] There are two different undergoer voice forms, here called undergoer voice 1 and undergoer voice 2. Verbal bases are lexically subcategorized for one of the two forms.[4] The following examples illustrate these different constructions. In (2), the verb *kaan* 'eat' is used in an actor voice construction, and in (3) in undergoer voice 1. Example (4) illustrates undergoer voice 2 with the verb *guru* 'learn'.

(2) *Aku mangaan sagin.*
aku moN-kaan sagin
1SG AV-eat banana
'I eat bananas.'

(3) *manuk tu kaan=na*
chicken DIST eat:UV1=3SG.GEN
'he eats the chicken' [chicken_eagle.275]

(4) *Tinga Inggris lau guru-i mangana ana.*
language English presently learn-UV2 child MED
'The child is learning English.'

Voice alternations in Totoli interact with various other grammatical and syntactic features. In the context of the present study, the most relevant features include

3. As common in the Austronesianist literature, we use the terms ACTOR and UNDERGOER as semantic macro-roles, i.e. generalized semantic relations, that subsume more agent-like roles like agents, experiencers, instruments, etc. under the label ACTOR, and more patient-like roles like patients, themes, locatives, etc. under the label UNDERGOER (cf., e.g. Foley & Van Valin 1984: 28ff; Van Valin & LaPolla 1997: 139ff).

4. In addition to the actor and undergoer voice, a third construction, the locative voice, is used when the subject is a locative argument (i.e. the place where an event takes place). This third voice is much more constrained in its uses than the other two voices and hence much less frequent (there were only 25 instances of locative voice in the corpus). Locative voice is not further discussed in this paper.

© 2022. John Benjamins Publishing Company
All rights reserved

mood, valency-increasing morphology, and relative and controlled complement clauses. Valency-increasing morphology, i.e. applicative and causative formations, are discussed in Section 4.1.2. All other features are briefly explained in the following paragraphs.

Totoli voice formations systematically occur in two moods: realis and non-realis. Realis mood denotes past events or situations that already exist and are still ongoing, while non-realis mood is used in reference to situations that do not (yet) obtain at the time of speaking. Non-realis mood is formally as well as semantically the unmarked member of the pair. The systematic relation between mood and voice marking is shown in the paradigm of formatives in Table 1.

Table 1. Dynamic voice and mood paradigms in Totoli

	NON-REALIS	REALIS
ACTOR VOICE	*mo-/moN-/mog-*	*no-/noN-/nog-*
UNDERGOER VOICE 1	*Ø*	*ni-*
UNDERGOER VOICE 2	*-i*	*ni- + -an*
LOCATIVE VOICE	*po-/poN-/pog- + -i*	*ni- + po-/poN-/pog- + -an*

Voice choice in general is not grammatically determined, i.e. there are no grammatical rules of the form 'in construction X the verb has to appear in actor voice'. However, in two types of subordinate clauses the voice of the predicate is conditioned by grammatical factors. In relative clauses, the head noun obligatorily functions as the subject of the relative clause and hence the voice of the verb has to be chosen accordingly. That is, a relativized actor triggers an actor voice construction and a relativized undergoer triggers an undergoer voice construction, as seen in Examples (5) and (6).

(5) *singgaian=na ttolu ia anu nog-uad-i isia itu=mo=ko itu*
friend=3SG.GEN three PRX REL AV.RLS-help-APPL2 3SG DIST=CPL=AND DIST
'three friends of his who helped him' [pearstory_23_AT.088]

(6) *bali mangana anu ala ia*
so little.child REL fetch:UV1 PRX
'so the children that were taken' [funeral_rites_TTL.182]

As for controlled complement clauses, there is a strong tendency for the controlled argument (i.e. the argument of the controlled verb that is necessarily coreferential with one of the arguments of the matrix verb) to function as the subject of the controlled verb, as in (7).

© 2022. John Benjamins Publishing Company
All rights reserved

(7) *I Epik nolinggo _____ mamakii tualina.*
i Epik no-linggo moN-paki-i tuali=na
HON PN ST.RLS-be.afraid AV-injure-APPL2 younger.sibling=3SG.GEN
'Epik was afraid to injure her younger brother.' (Riesberg 2014: 38)

Unlike in the case of relative clauses, however, the restrictions on subject choice in controlled clauses are not without exceptions. Some speakers also allow the controlled argument to occur in non-subject function as in (8) (see Riesberg 2014: 37–43 for further details and discussion).

(8) *Inangku mogole Ismail aku monurungi _____.*
inang=ku mo-gole Ismail aku moN-turung-i
mother=1SG.GEN AV-ask.for PN 1SG AV-help-APPL2
'My mother asks Ismail for me to help (him).' (Riesberg 2014: 41)

Inasmuch as subject choice in complement constructions is restricted to controlled arguments, voice choice for the complement verb is grammatically determined. As in all of our corpus examples the controlled argument does in fact function as the subject, we have excluded the voice of controlled verbs from our investigation to avoid a bias based on grammatical restrictions.

2.2 Corpus

For the present study we annotated a total of 2 hours 50 minutes of spoken texts recorded during various field trips between 2006 and 2018, with the exception of one recording which dates from 1989.[5] The corpus consists of 27 texts involving a total of 53 different speakers, 26 female and 27 male, mainly adults (only one text involves children and teenagers), all of whom live in bilingual Totoli/Indonesian-speaking households.

The texts amount to 16,272 words and consist of 6,745 intonation units, as defined in Himmelmann et al. (2018). Hesitations, false starts, non-linguistic signals and inaudible segments are marked in the transcripts but excluded from all counts. From this data, only those symmetrical voice constructions where a choice between actor voice or undergoer voice is possible were included in the statistical analysis. This excludes, first, non-voice-marked predicates and intransitive clauses. Secondly, we excluded relative clauses and complement clauses, since voice choice in these constructions is syntactically determined (see previous subsection). Lastly, we also left out clauses in Indonesian. Importantly, however, all

5. This latter text was recorded by Nikolaus P. Himmelmann during a survey of the Tomini-Tolitoli area between August 1988 and January 1989. All other texts were collected during two recent documentation and research projects (see Footnote 1 and Acknowledgements).

© 2022. John Benjamins Publishing Company
All rights reserved

referents that occur in the omitted clauses have been annotated and considered as possible antecedents.

All in all, a total of 962 transitive clauses thus entered the analysis, of which 690 (72%) are undergoer voice constructions and 272 (28%) are actor voice constructions. Comparing this ratio to the figures reported for corpora from other western Austronesian languages (listed in Table 2), we find substantial differences, with some corpora showing a bias towards the actor voice and others being strongly biased towards the undergoer voice, as is the case in our Totoli corpus. Note, however, that comparing voice distributions across studies is hampered by the fact that the corpora used in these studies are not directly comparable. Thus, for example, the corpus used by Pastika (1999) consists of spoken and written materials in almost equal proportions. Furthermore, the spoken part of the corpus was contributed to a large extent by skilled story tellers using a fairly formal speaking style, with few hesitations and a substantial number of complex constructions. Wouk (1999: 94) reports for her corpus of conversational Sasak (Ngeno-Ngené) a ratio of 47% actor voices and 53% undergoer voices (345 transitive clauses in total). Asikin-Garmager's corpus of the same Sasak variety, consisting only of narratives (and almost exclusively of re-tellings of the Frog Stories) contains 66% actor voices and 34% undergoer voices (324 transitive clauses) (Asikin-Garmager 2017: 238).[6] Consequently, it would be a mistake to attribute the differences seen in Table 2 directly to the languages. Nevertheless, it is also clear that there are differences in the usage conditions for voice alternations across western Austronesian symmetrical voice languages, a topic, however, that we cannot pursue further here.

Corpora not only show differences in the ratios of actor and undergoer voices, but also in the numbers of transitive constructions. For instance, McDonnell's corpus (2016) contains 899 transitive symmetrical voice constructions (excluding all clauses in which voice choice is predetermined) in 80 minutes of recording. Our corpus has 962 transitive clauses in 170 minutes, which indicates that transitive voice constructions occur considerably less frequently in the Totoli corpus than in the Besemah corpus. In all likelihood, such differences are mostly due to the text types represented in the corpora and not so much to the grammars of the languages.

6. Note that Sasak differs from the other western Austronesian languages discussed in this paper in that the functional difference between bare verbs forms and verbs marked by the nasal prefix (i.e. the Proto-Austronesian actor voice prefix) is neutralized in certain contexts in some varieties, or even completely given up in others. It is thus not straightforward to decide to what extent Sasak actually fits the definition of a symmetrical voice language given in Riesberg (2014: 10). See Wouk (1999), Arka (2009), and Kroeger & Riesberg (forthcoming) for further discussion.

© 2022. John Benjamins Publishing Company
All rights reserved

Table 2. Ratios of actor voice and undergoer voice in some Austronesian symmetrical voice languages

	% AV	% UV	Trans. clauses	Reference
Balinese	62	38	1851 (+ 8% PASS)	Pastika (1999: 62)
Besemah	56	44	899	McDonnell (2016)
Pendau	45	55	443	Quick (2005: 235)
Tondano	30	70	314	Brickell (2014)
Totoli	28	72	962	this study
Classical Malay	27	73	115	Cumming (1991: 85)

The most significant difference between the Totoli corpus and corpora used in previous studies lies in the variety of text genres included. We have made a major effort to ensure that the corpus is diverse along two major parameters (in addition to representing the speech of a largish number of speakers, as opposed to the handful or less of speakers represented in other corpora). On the one hand, the recordings included differ in their **degree of interactivity**, ranging from almost pure monologue to highly interactive everyday conversation involving up to eight speakers. All recordings were made in a Totoli-speaking environment where the speaker received feedback from at least one other native speaker. Consequently, the distinction between monologues and dialogues made in Table 3 and later on in Section 6 is somewhat arbitrary, as it imposes a boundary on a continuum. Monologues are essentially defined by the fact that the overall speech event was set up as one person telling a story or explaining a procedure. In all but one instance, the speakers did not prepare their speech beforehand but rather created the story or explanation on the spot. Hence, monologues in this corpus are also fairly spontaneous and often involve a fair number of clarification questions and comments by the audience. The one exception is the recording of a folk tale from 1989, which is based on a written version produced by the speaker beforehand, but which considerably diverges from the script after the first few lines.

The second parameter of variation pertains to the genre or text type, essentially defined by the topic of the recorded speech event. As summarized in Table 3, dialogues include everyday conversations (people talking about crying babies, their German visitors, plans for tomorrow, neighbourhood gossip, etc.), personal experiences, procedural texts, and task-oriented dialogues, which include Space Games (Levinson et al. 1992) and the recounting and discussing of video stimuli. As for monologues, we included folk tales, personal experiences, procedural texts, and (monological) retellings of the Pear Film (Chafe 1980). The retellings of the

© 2022. John Benjamins Publishing Company
All rights reserved

Table 3. Overview of the Totoli corpus

Dialogue	Texts	Hours	IUs	Tr. clauses	Monologue	Texts	Hours	IUs	Tr. clauses
conversation	5	0:22:27	1188	119	folk tales	6	0:35:25	1336	193
personal experience	1	0:16:12	486	97	personal experience	2	0:11:34	435	25
procedural	2	0:14:32	878	127	procedural	2	0:20:12	578	147
task-oriented	4	0:32:53	1297	135	retelling of Pear Film	5	0:16:49	547	119
Totals	**12**	**1:26:04**	**3849**	**478**		**15**	**1:24:00**	**2896**	**484**

Pear Film differ from the other monologue types in that they deal with a topic never discussed before.

2.3 Corpus annotation

The data in our corpus have a morpheme-by-morpheme gloss, to which three further layers of annotation were added: one each for GRAID, RefIND and RefLex. GRAID (*Grammatical Relations and Animacy in Discourse*) is an annotation convention for coding the syntactic function, morphological form, and animacy of referential expressions. It was developed by Haig & Schnell (2015) with the purpose of facilitating quantitative crosslinguistic investigations of natural discourse.

In accordance with the GRAID annotation guidelines, the annotations in our corpus register:

- The form of referential expressions (i.e. argument realization): The most central distinction in GRAID is made between noun phrases headed by nouns, pronouns, and zero arguments. Referents glossed as zero arguments include only cases where the argument is required by the valency of the verb and is also expressible in an overt form.
- Semantic and pragmatic properties of referential expressions: The two semantic parameters annotated are person and animacy. Regarding animacy, a four-way distinction is made between humans, anthropomorphized referents (i.e. animals or plants with human-like characteristics and abilities as they often occur in traditional folk tales), animals, and non-animate referents.
- Grammatical relations: Haig & Schnell (2015) make use of the three-fold classification of syntactic functions of core arguments widely used in typological studies and distinguish between the single argument of a one-place predicate, and the more actor-like and the more undergoer-like argument of a tran-

© 2022. John Benjamins Publishing Company
All rights reserved

sitive predicate. However, given the symmetrical voice system of Totoli, in which the grammatical subject and object of a transitive clause may align both with the agent and the patient roles, we adapted the GRAID scheme as follows. In our corpus, core arguments of transitive sentences are coded for their semantic role (actor or undergoer) and for the voice of the predicate they are arguments of. From the combination of these two labels we can infer their syntactic functions. The actor of an actor voice and the undergoer of an undergoer voice predicate function as subject arguments whereas the actor of an undergoer voice and the undergoer of an actor voice predicate function as non-subject arguments. The sole argument of an intransitive predicate and non-core arguments (goals, locatives, obliques, dislocated topics, and circumstantials) are annotated as such.

- Predicates: Verbal predicates are glossed for voice and lexical aspect (dynamic vs. stative).
- Clause boundaries: GRAID annotations distinguish dependent from independent clauses. Among dependent clauses, a further distinction is made between relative, complement and adverbial clauses.

RefIND (*Referent Indexing in Natural-language Discourse*; Schiborr et al. 2018) is an annotation convention used to track discourse referents. Each referent receives a numeric index, which is annotated throughout the text at each mention of the referent, irrespective of the lexical item used to refer to it and of whether it is an overt or a non-overt mention. The main challenge posed by this annotation scheme is to decide what is considered a discourse referent and what is not. For our corpus, we followed Schiborr et al.'s (2018) guidelines and annotated all identifiable and trackable entities, which include specific physical real-word entities, non-specific entities if they are taken up again in subsequent discourse, and generic classes of referents.

RefIND is accompanied by a simplified version of the RefLex annotation conventions (Baumann & Riester 2012, Riester & Baumann 2017), which tags referents according to their information status.[7] Our simplified annotation scheme

7. The central idea of Baumann & Riester's (2012) Ref(erential)Lex(ical) scheme is that information status should be analyzed at two separate levels. In the referential domain, the status of a constituent depends on whether the referring expression has a co-referential antecedent. By contrast, in the lexical domain, expressions are labeled as *l-given* only if the lexical expression itself was used in the previous discourse, regardless of whether it is used in reference to the same or a different discourse referent. The simplified version of RefLex used with RefIND here marks information status only at the referential level. Moreover, Baumann & Riester 2012 propose a much more fine-grained classification scheme for referential and lexical expressions, only four labels of which are applied here.

© 2022. John Benjamins Publishing Company
All rights reserved

involves a four-way distinction between **given, new, unused**, and **bridging**. The status **given** is assigned to all referents with a RefIND index that have already been mentioned in the previous discourse. First mentions of referents with a RefIND index are labelled as **new**, except for globally known referents (labelled as **unused**) and referents inferable from context (labelled as **bridging**).

The combination of RefLex and RefIND is used in the statistical analyses in order to evaluate the influence of discourse-related properties of arguments on voice choice, including, for example, TOPICALITY, ACTIVATION STATE, and TRACKING function.

2.4 Summary and note on statistics model

Our corpus contains 962 clauses out of a total of 27 texts of diverse interactivity and text types. The clauses included are all transitive and voice-marked (excluding locative voice), and include both main clauses and adverbial subordinated clauses but not relative or controlled complement clauses. The referents occurring in these clauses and, importantly, also the referents of the excluded clauses are coded for grammatical relations, form of expression (lexical, pronoun, zero), and animacy on the GRAID level. They are also numerically indexed on the RefIND layer, and are tagged for information status on the RefLex tier.

For each of the variables investigated in the following four sections, a generalized linear mixed-effects model was fitted, using the function glmer() from the package *lme4* (Version 1.1.17; Bates et al. 2015) in R (Version 3.4.0; R Core Team 2017) considering speaker as a random effect. All bar graphs that show the distribution of actor voice and undergoer voice in relation to the factors we investigate include a black dashed line. This line indicates the overall distribution of actor voice and undergoer voice in the set of transitive clauses represented in the graph (i.e. it indicates the 28% actor voice to 72% undergoer voice ratio reported in Table 2 when all 962 transitive clauses are taken into account, or the respective distribution of actor voice and undergoer voice in a particular subset of this total).

3. Argument-related factors

In this section, we look at properties that pertain to the two core arguments of a transitive verb. These properties can be reference related or they can be discourse related, that is, they are either inherent to the referent (such as ANIMACY), or they only arise due to their actual use in discourse (such as TOPICALITY). We discuss the two types of properties separately in the following two subsections.

© 2022. John Benjamins Publishing Company
All rights reserved

3.1 Argument-inherent properties: Animacy and humanness

The question whether there is a correlation between voice choice and the ANIMACY of the arguments in a transitive clause has often only been addressed indirectly. Cooreman, for example, found for Chamorro that animate referents tend to be more topical (Cooreman 1983: 453f, see Section 3.2 below for a more detailed discussion of the notion of topicality), and that, in turn, topicality plays a role in voice selection (Cooreman 1983: 463ff). However, no explicit counts are given that would show a direct correlation between ANIMACY and voice. There are, as far as we are aware, three studies that do explicitly investigate this interrelation. For Besemah, McDonnell (2016: 210f) finds that ANIMACY is not a statistically significant factor for voice choice, even though undergoer voice is slightly more frequent when the undergoer is animate. Similarly, Asikin-Garmager (2017: 245f) finds that animacy has no statistically significant effect on voice choice in his corpus of Sasak (Ngeno-Ngené) narratives, even though in a previously conducted production experiment he sees a significant increase in the use of undergoer voices if the agent is inanimate and the undergoer is animate, and a decrease of undergoer voice in the reverse condition (Asikin-Garmager 2017: 174). Cumming (1991), in her study of Classical Malay texts, distinguishes between human and non-human referents and, likewise, finds that this distinction does not play a significant role in voice choice. One could note in passing that all actor arguments in her corpus are human and that when the undergoer is human, there is actually a slight preference for actor voice (Cumming 1991: 131).

In our study, we tested for both the animate-inanimate and the human–non-human distinction. In the latter, the human category also comprised anthropomorphic animals, i.e. animals in traditional folk stories that act like humans in that they can talk and experience adventures as human protagonists would. The prediction is that if only one argument is animate or human, then this argument should be linked to the subject relation; hence actor voice when this argument functions as actor, and undergoer voice when it functions as undergoer. In case both arguments are equal with regard to animacy or HUMANNESS, no strong preference for either voice is predicted.

Figure 1 shows the results for HUMANNESS, none of which fits the prediction. By far the most frequent configuration is the one in which a human actor is acting on a non-human undergoer (second column). With 29% actor voices and 71% undergoer voices, this configuration also almost exactly reflects the distribution of actor voice and undergoer voice in the overall corpus (28% actor voice vs. 72% undergoer voice, see Table 2 in Section 2). When both actor and undergoer are human, undergoer voice is more frequent. On the other hand, when the actor is non-human (last two columns), an above-average number of actor voices is

© 2022. John Benjamins Publishing Company
All rights reserved

attested, though, again, we see slightly more instances of the undergoer voice if the undergoer is human. The configuration in which a non-human actor acts on a non-human undergoer evokes a ratio of actor voice to undergoer voice (67% actor voice vs. 33% undergoer voice) that is almost the reverse of the overall distribution in the corpus. These include examples of the type *the west wind can destroy the Durian blossom*. Note, though, that for both combinations that involve a non-human actor the overall numbers are small.

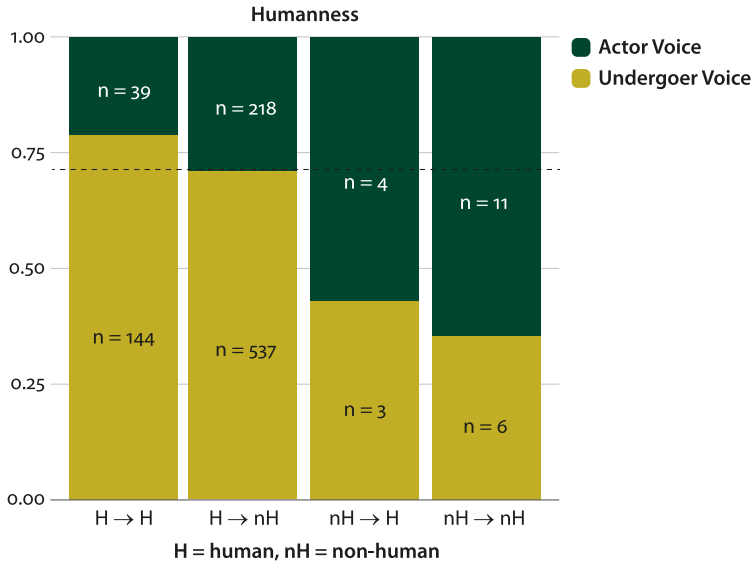

Figure 1. Humanness and voice in Totoli. Number of transitive clauses = 962

Looking at ANIMACY (Figure 2), we see a similar picture, again contrary to the prediction. The largest number of all transitive clauses in our corpus describe events in which an animate actor acts on an inanimate undergoer. The tendency for animate undergoers (where the actor is also animate) to more strongly favour undergoer voice is found here as well (25% actor voice vs. 75% undergoer voice), but it is slightly less pronounced than in the human–human distinction (21% actor voice vs. 79% undergoer voice). On the other hand, the reversal of the actor voice-undergoer voice ratio when the actor argument is inanimate is even more pronounced than when the actor is non-human. However, here the numbers are even smaller, and with a total of fourteen tokens for the configurations **inanimate actor acts on animate undergoer** and **inanimate actor acts on inanimate undergoer** (last two columns), it seems questionable to make any claims in this regard. Still, generalized linear mixed-effects model analyses revealed a statistically significant association between voice and both humanness and animacy (see Supplement for details).

© 2022. John Benjamins Publishing Company
All rights reserved

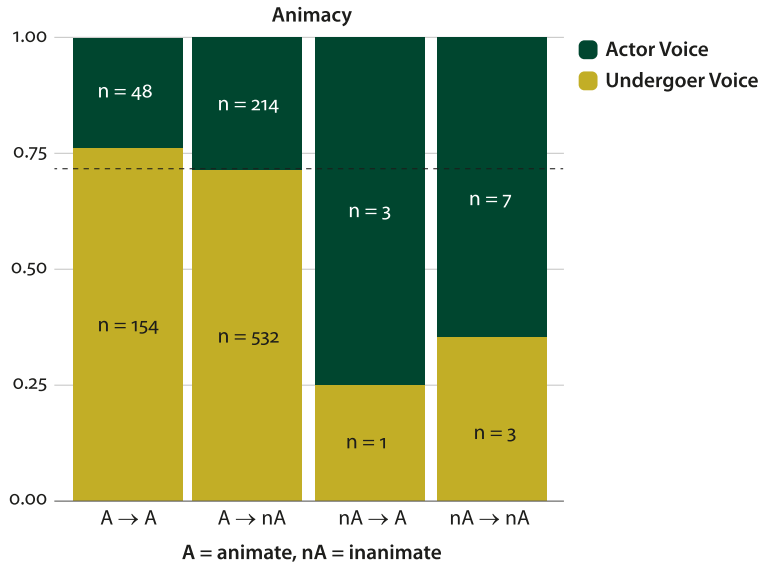

Figure 2. Animacy and voice in Totoli. Number of transitive clauses = 962

3.2 Discourse-related properties: Activation state, topicality, discourse reference function, generalizability

Unlike the factors discussed in the previous section (HUMANNESS and ANIMACY), the properties discussed in this section are not inherent to the referent referred to by a given nominal expression, but rather arise through their actual use in discourse. In the following four subsections, we investigate whether the TOPICALITY, the ACTIVATION STATE, the REFERENCE FUNCTION, or the GENERALIZABILITY of arguments affects the speakers' choice of voice in Totoli.

3.2.1 Topicality

TOPICALITY is probably the most widely investigated concept in Austronesian discourse studies. Following Givón (1983), there are a number of works that investigate topic continuity in discourse, such as Cooreman (1983, 1988) for Chamorro, Pastika (1999) for Balinese, Payne (1994) for Cebuano, Brainard (1994) for Karao, Quick (2005) for Pendau, Wouk (1996) for Jakarta Indonesian, or Wouk (1999) and Asikin-Garmager (2017) for Sasak. In these studies, TOPICALITY is determined by two discourse measures: REFERENTIAL DISTANCE and TOPIC PERSISTENCE.[8]

8. In the introduction to the widely quoted edited volume on topic continuity (1983), Givón in fact proposes three measurements: REFERENTIAL DISTANCE, POTENTIAL INTERFERENCES, and

© 2022. John Benjamins Publishing Company
All rights reserved

REFERENTIAL DISTANCE (RD) measures the degree of continuity of a given NP by looking at the previous discourse environment, i.e. by "looking to the left". In the literature, REFERENTIAL DISTANCE has been measured in three different ways: First, as originally suggested by Givón (1983), by counting the number of clauses between a given referential expression and its last mention (cp. also Cooreman 1983, Payne 1994). The maximum value is arbitrarily set to 20 clauses, and "referential indefinite" NPs and first-mentioned NPs are assigned this maximum value of 20 (Givón 1983: 13). Second, by applying either a three-way distinction (1, 2–3, >3; cp. Givón 1994, Brainard 1994, Wouk 1999) or a two-way distinction (1–3, >3; cp. Quick 2005). And, finally, by looking at the relative distance from the two arguments in a transitive clause to their previous mentions (cp. Dryer 1994, Quick 2005, Asikin-Garmager 2017). That is, authors employing this third method measured whether the REFERENTIAL DISTANCE of the actor argument is shorter than the REFERENTIAL DISTANCE of the patient argument, or the other way round. For the measure of REFERENTIAL DISTANCE in Totoli, we applied the third method. All approaches assume that a shorter distance between a referential expression and its previous mention equals a higher degree of TOPICALITY.

For the above-mentioned languages, we can observe similar tendencies, even if the languages differ with respect to certain details. Thus, for the corpora of Jakarta Indonesian (Wouk 1996), Balinese (Pastika 1999), and Pendau (Quick 2005), for example, it holds that the degree of topicality of actor and undergoer in undergoer voice is roughly the same (both are highly topical), but that in actor voice the two arguments differ dramatically in that the undergoer has a much lower degree of topicality than the actor. The differences are of the following type: While in Balinese, Pendau, and Sasak the undergoer-voice actor is slightly more topical than the undergoer (cp. Pastika 1999: 199, Quick 2005: 230, Asikin-Garmager 2017: 247), the opposite is the case in Jakarta Indonesian (Wouk 1996: 379). Furthermore, in the Pendau and Jakarta Indonesian corpora, the undergoer-voice actor is (slightly) more topical than the actor-voice actor, while in Balinese the actor-voice actor shows slightly higher topicality than the undergoer-voice actor (Pastika 1999: 199). It is difficult to say to what extent such differences arise from different ways of coding and counting the data, or whether they in fact reflect differences in the corpora or, even more generally, in the usage conditions established within the different speech communities.

The Totoli corpus adheres to the overall trend. Just like in the corpora of Balinese, Pendau and Jakarta Indonesian, we find only little difference in the average REFERENTIAL DISTANCE of actors and undergoers in the undergoer voice, but

PERSISTENCE. As subsequent studies only applied REFERENTIAL DISTANCE and PERSISTENCE, we, too, do not discuss POTENTIAL INTERFERENCE here.

© 2022. John Benjamins Publishing Company
All rights reserved

considerable variation in the actor voice, where actors are much more topical than undergoers. In this analysis, Totoli patterns with Balinese and Pendau in that in the undergoer voice, the actor is more topical than the undergoer. Furthermore, again like Balinese but unlike Pendau and Jakarta Indonesian, the Totoli actor-voice actor is more topical than the undergoer-voice actor. Table 4 provides the average REFERENTIAL DISTANCE (measured in clauses, as originally proposed by Givón 1983) for Totoli and Jakarta Indonesian (JI), as provided in Wouk (1996: 379).

Table 4. Average RD of actors and undergoers in Totoli and Jakarta Indonesian

	Totoli	**JI**
AV actor	2.67	5.08
AV undergoer	4.39	13.7
UV actor	3.47	5.04
UV undergoer	3.84	4.66

Figure 3 shows the relative REFERENTIAL DISTANCE of actor (ACT) and undergoer (UG) arguments and voice choice in Totoli on the basis of a dataset that only considers clauses in which both actor and undergoer have been previously mentioned. The total number of transitive clauses in this analysis thus amounts to only 258. Following previous studies, the cut-off point is 20 clauses (i.e. previous mentions further away than 20 clauses are not considered). We see that for most transitive clauses in this dataset, the most frequent configuration is for actor and undergoer to have the same REFERENTIAL DISTANCE. If the actor has a lower REFERENTIAL DISTANCE than the undergoer, actor voice is used slightly more often. And the other way round, if the undergoer has a lower REFERENTIAL DISTANCE than the actor, we observe a slightly higher use of undergoer voices. A generalized linear mixed-effects model revealed a statistically significant association between voice and referential distance (see Supplement for details). Figure 4 (generated using the counts provided in Quick 2005) shows the same tendency for the Pendau dataset, where the preference for actor voice when the actor is more topical than the undergoer is more pronounced than in the Totoli corpus, and where equal REFERENTIAL DISTANCE of actor and undergoer seems to trigger a higher likelihood of undergoer voice. In the Sasak (Ngeno-Ngené) corpus, on the other hand, Asikin-Garmager only finds an increased use of undergoer voice of when the undergoer is more topical than the actor. A higher topicality of the actor over the undergoer shows no effect, i.e. no preference for actor voice (Asikin-Garmager 2017: 250f).

© 2022. John Benjamins Publishing Company
All rights reserved

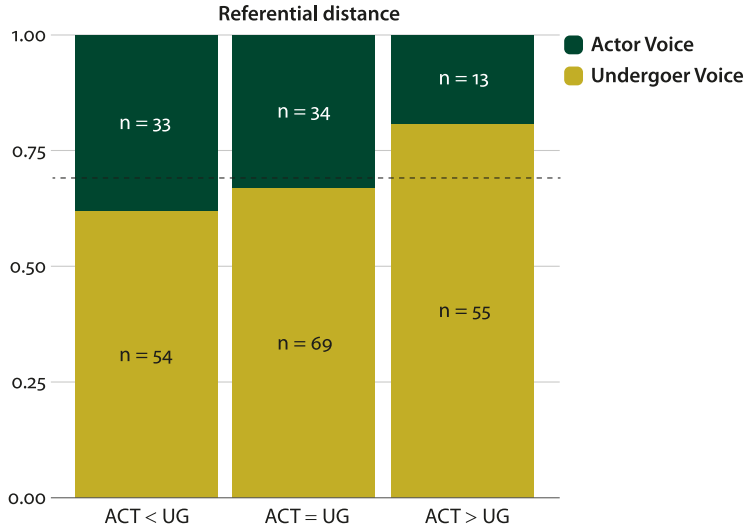

Figure 3. RD (up to 20 clauses back) and voice in Totoli. Number of transitive clauses = 258

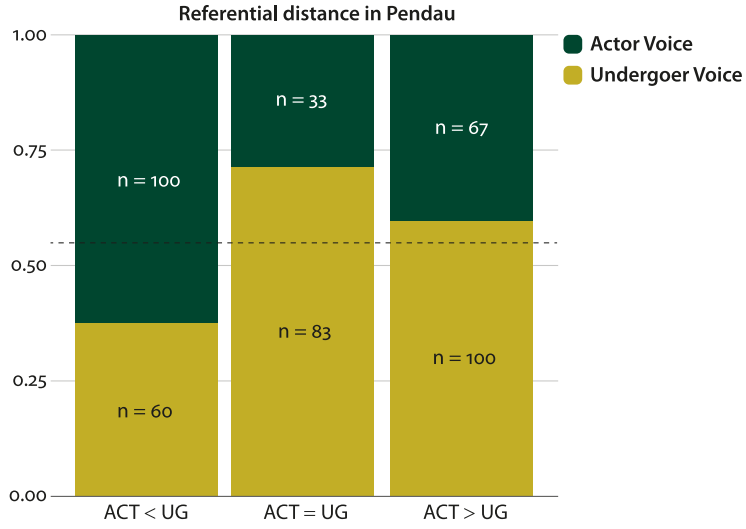

Figure 4. RD and voice in Pendau from Quick (2005: 235). Number of transitive clauses = 443

TOPIC PERSISTENCE (TP) counts all clauses following a referential expression (i.e. 'looking to the right') that contain the same referent. Just as with REFERENTIAL DISTANCE, previous studies differ in the exact measures that have been

© 2022. John Benjamins Publishing Company
All rights reserved

applied. Earlier works have counted single clauses, with no maximum value being assigned (i.e. no limit has been applied to the right; e.g. Givón 1983, Cooreman 1983). Later works have restricted their counts to a maximum number of ten clauses, and thus to a maximum assigned value of ten (e.g. Givón 1994, Payne 1994). Others have applied a similar two-way distinction to the one described for REFERENTIAL DISTANCE above, i.e. distinguishing between a persistence of two or less clauses and a persistence of three or more clauses (e.g. Dryer 1994, Pastika 1999, Wouk 1999). For our analysis, we counted mentions of referents up to ten clauses following the target expression. Despite the variation in the details of measuring TOPIC PERSISTENCE, there is broad consensus that the longer a referent keeps being mentioned within a given stretch of discourse, the more topical it is.

In the Balinese corpus, measuring TOPIC PERSISTENCE shows pretty much the same results as when measuring REFERENTIAL DISTANCE: both undergoer-voice actors and undergoers have a high persistence (i.e. are highly topical). Actor-voice actors, on the other hand, are much more persistent than actor-voice undergoers (Pastika 1999: 199). In the Jakarta Indonesian data, actor-voice actors have a higher persistence than actor-voice undergoers, and undergoer-voice undergoers have a higher persistence than undergoer-voice actors. It is thus the subject arguments that show the highest degree of topicality, with undergoer-voice undergoers being slightly more persistent than actor-voice actors (Wouk 1996: 380). In the Totoli corpus, on the other hand, it is the actor-voice arguments that exhibit an overall higher persistence than the undergoer-voice arguments, and, like in Jakarta Indonesian, undergoer-voice undergoers are more persistent than undergoer-voice actors. Table 5 compares the average TOPIC PERSISTENCE values in the Totoli and spoken Jakarta Indonesian corpora.

Table 5. Average TP of actors and undergoers in Totoli and Jakarta Indonesian

	Totoli	**JI**
AV actor	1.75	1.35
AV undergoer	1.58	0.7
UV actor	1.36	0.81
UV undergoer	1.44	1.5

Figure 5 shows that relative TOPIC PERSISTENCE has practically no influence on voice choice in the Totoli corpus. Neither the undergoer being more persistent than the actor, nor equal topicality of actor and undergoer in terms of persistence has any effect on voice distribution. When the actor has higher TOPIC PERSISTENCE there seems to be a slight preference for actor voice, but this effect was

© 2022. John Benjamins Publishing Company
All rights reserved

not found to be statistically significant in a generalized linear mixed-effects model analysis (see Supplement for details).

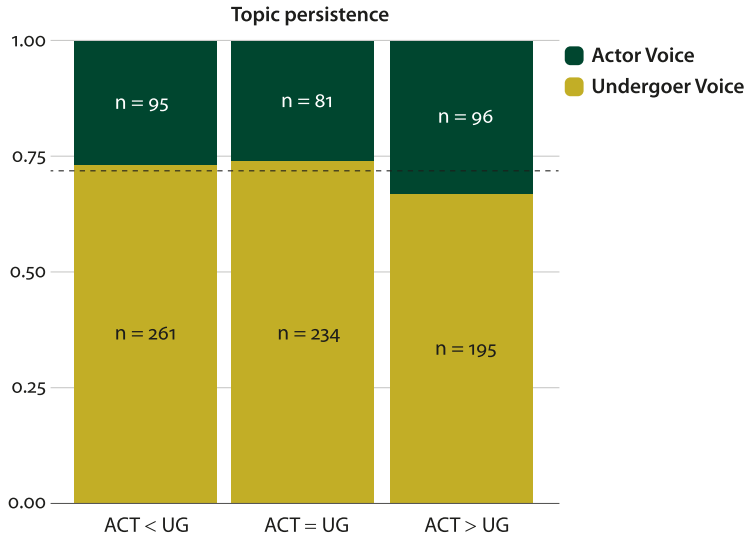

Figure 5. TP and voice in Totoli. Number of transitive clauses = 962

3.2.2 Activation state and argument realization

The ACTIVATION STATE of referents in discourse has been an important concept in the study of reference for a long time, and there is a lot of research that shows how the accessibility of a referent determines the speaker's choice in the form of referring expressions. Different approaches have been proposed concerning the gradience of activation. Early works (e.g. Chafe 1976, Prince 1981) distinguish three activation statuses – given, activated and new. Other authors have proposed more fine-grained distinctions, such as those posited in the givenness hierarchy established by Gundel et al. (1993), which consists of six statuses, each of which is assumed to be "a necessary and sufficient condition for the appropriate use of a different form or forms" (Gundel et al. 1993: 275). However, Gundel et al. show in their comparative study of English, Chinese, Japanese, Russian and Spanish that not all statuses are relevant in all languages. The basic expectation is that the more activated (given) argument is the preferred choice for the subject relation. Similarly, the argument given non-lexical expression should be the preferred subject in case the other argument is given lexical expression.

As described in Section 2.3, we applied the RefIND annotation scheme to our Totoli corpus. Hence, all expressions referring to identifiable and trackable entities in our corpus were assigned a unique identifier and were annotated with

© 2022. John Benjamins Publishing Company
All rights reserved

the labels **given, new, bridging** or **unused**. In our analysis of the ACTIVATION STATE of referents, we included **bridging** and **unused** referents under the status **given**. Figure 6 shows how given and new referring expressions are distributed over actor and undergoer voices. All clauses involving arguments that have not been given a RefIND label in the first place, i.e. non-referential expressions, are excluded from this analysis. Figure 6 is thus based on a reduced number of 432 transitive clauses only, rather than on the 962 transitive clauses that make up our full corpus.

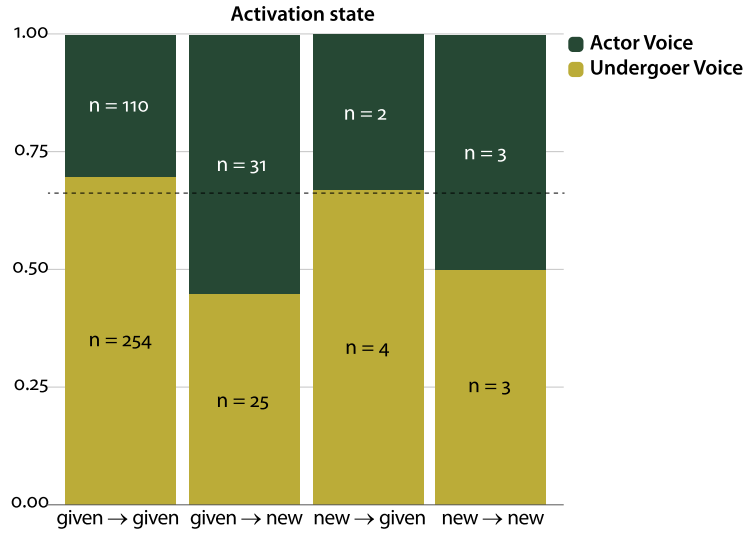

Figure 6. Activation state and voice in Totoli. Number of transitive clauses = 432

In our corpus, the ACTIVATION STATE of the actor does not seem to have any influence on voice choice – compare the first with the third column, and the second with the fourth column. However, when the undergoer argument is new, actor voice is more frequent (irrespective of whether the actor is new or given). This effect was found to be statistically significant in a generalized linear mixed-effects model analysis (see Supplement for details), though we note that new undergoer arguments in transitive clauses are relatively rare (62 clauses) compared to transitive clauses with given undergoers (370 clauses). This effect is even more pronounced in the Besemah corpus (McDonnell 2016: 221), where a much larger proportion of actor voices (93%) occurs when the actor is given and the undergoer is new. However, for Besemah data we also see a preference for undergoer voice when the undergoer is given. This preference is more pronounced when the actor is new (18% actor voice vs. 82% undergoer voice), but it still persists when both actor and undergoer are given (48% actor voice vs. 52% under-

© 2022. John Benjamins Publishing Company
All rights reserved

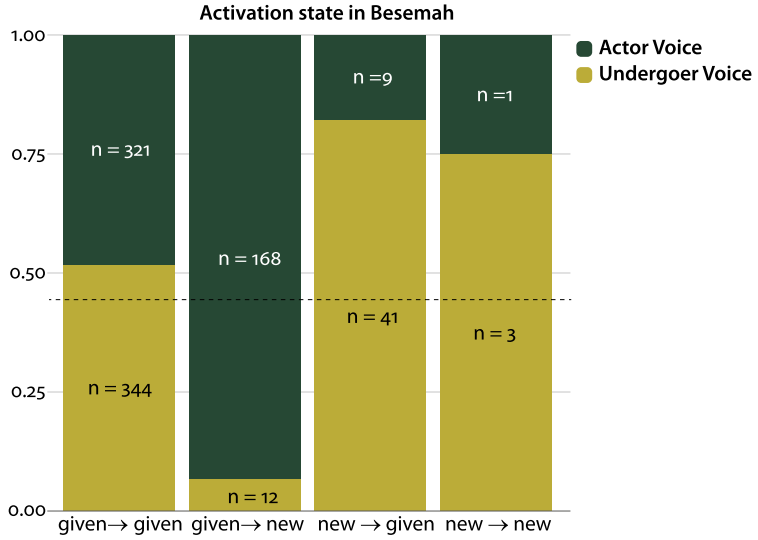

Figure 7. Activation state and voice in Besemah (McDonnell 2016: 221). Number of transitive clauses = 899

goer voice).[9] In both the Totoli and the Besemah corpus, however, configurations involving a new actor and a given undergoer and those involving two new participants are very rare.

As mentioned at the beginning of this section, it is well known that the ACTIVATION STATE of a particular referent determines the form of the referring expression which is appropriate in a particular situation. In this sense, ARGUMENT REALIZATION can be taken as a proxy for the ACTIVATION STATE, and we therefore investigated the interrelation between ARGUMENT REALIZATION and voice. Arguments realized as full NPs or free relative clauses, as well as clausal arguments (complement clauses) are counted as lexical realizations, and pronouns and zeros (as well as demonstratives and the filler *anu* 'thing, whatchamacallit') as non-lexical, assuming that an argument realized as a pronoun or as zero is arguably more activated than an argument realized as a full NP.[10] The reason we are using the lexical–non-lexical distinction as a further measure to determine the ACTIVATION STATE of a referent is twofold. First, it is conceivable that it is slightly more accurate than the **given–new** distinction described above. This is because any re-mention of a referent is counted as **given**, no matter how far away its last mention

9. The overall voice distribution in McDonnell's corpus is 56% actor voice vs. 44% undergoer voice (see Section 2, Table 2).

10. The decision to count demonstratives and fillers as non-lexical is rather problematic in this respect, but note that this does not considerably affect the overall tendencies.

© 2022. John Benjamins Publishing Company
All rights reserved

might be. It is unlikely, however, that a referent that has not been mentioned for a very long time will be taken up again by a pronoun. In this sense, the lexical–non-lexical distinction probably reflects the ACTIVATION STATE of a referent more accurately than the categorical distinction between **given** and **new**. Second, and this is a purely technical reason, the lexical–non-lexical distinction can be applied to the whole corpus, i.e. to all of our 962 transitive clauses, making the results of a statistical analysis more robust.

Despite the differences just mentioned, we find that the overall tendencies for ARGUMENT REALIZATION (lexical vs. non-lexical) and voice choice mirror the patterns observed in Figure 6 above. As Figure 8 shows, there is a tendency for the actor voice to be used more often when the undergoer is lexical (i.e. corresponding to **new** in Figure 6), and for the undergoer voice to be used more often (relative to the overall distribution) when the undergoer is non-lexical. A generalized linear mixed-effects model confirmed the statistical significance of this association (see Supplement for details). This association is strongest when the actor argument, too, is non-lexical (which amounts to 89% of all tokens, and corresponds to the configuration **given–given** in Figure 6). Importantly, however, in the two configurations where the values for the two arguments differ – one being lexical, the other being non-lexical – the proportions of actor and undergoer voice are almost identical (shown in columns two and three in Figure 8). In this regard, the numbers do in fact differ somewhat between Figure 6 and Figure 8.

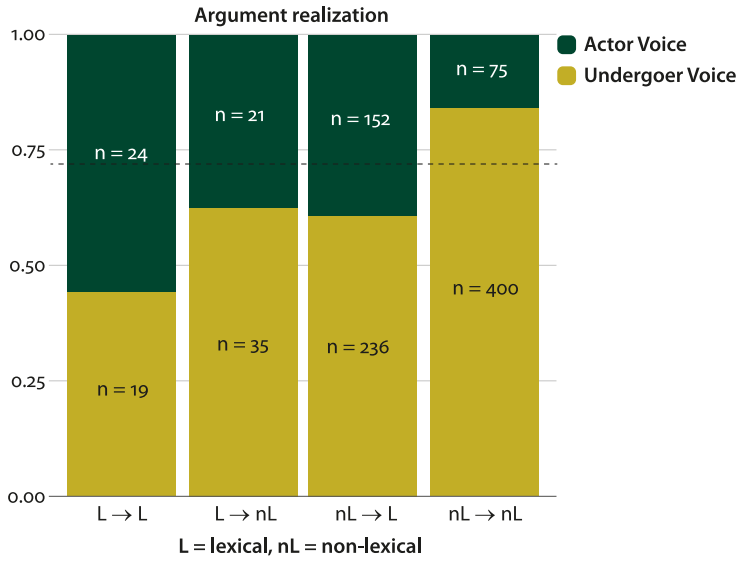

Figure 8. Argument realization and voice in Totoli. Number of transitive clauses = 962

© 2022. John Benjamins Publishing Company
All rights reserved

3.2.3 Tracking use

Referring expressions may have a TRACKING function; that is, they may help the listener to keep track of referents throughout the discourse. Otherwise, they are non-tracking (as all non-referential nominal expressions are by default). In our corpus, a referent is considered to be **tracking** if it is mentioned more than once (i.e. if it has been annotated with a RefIND/RefLex label and is mentioned at least once more after its first mention). All other argument expressions are considered **non-tracking**. The TRACKING distinction is related to the measure of TOPIC PERSISTENCE (Section 3.2.1) in that non-tracking expressions are by definition non-persistent. However, it differs from TOPIC PERSISTENCE in two ways. First, no limit is set on the number of clauses that occur between two mentions of a referent. Second, the number of times a referent is mentioned is not relevant beyond the fact that it is mentioned at least twice. Tracked arguments are expected to be preferred as subjects when in construction with a non-tracked argument.

Figure 9 shows that the most frequent configuration in the transitive clauses of our corpus is that both actor and undergoer are tracking. It further shows that it is the TRACKING function of the actor argument, but not of the undergoer that has an impact on voice choice, albeit a fairly minor one: if the actor is tracking, there is a slightly higher chance of actor voice. If the actor is non-tracking, undergoer voice is somewhat more frequent. The association between tracking status of the actor and voice choice was found to be statistically significant in a generalized linear mixed-effects model analysis. No such association is found between the tracking status of the undergoer and voice choice (see Supplement for details).

Again we can compare our Totoli results with McDonnell's findings for Besemah, illustrated in Figure 10 (cp. McDonnell 2016: 223). Here, differences between the two languages (or rather, corpora) are quite striking. Most importantly, there is a strong preference for actor voice when the actor is tracking and the undergoer is non-tracking, and a strong preference for undergoer voice in the reverse case. In the Besemah corpus it thus seems to be the case that trackable arguments are more likely to be selected as subjects, as to be expected if voice choice is related to subject choice. This is clearly not the case in Totoli. This difference is partially due to differences in coding. McDonnell, for example, coded bridged arguments also as tracking (2016: 222). But it may in fact reflect a genuine difference in the usage conditions for the two voices in the two languages. Note, however, that the tracking difference does not explain voice choice in those cases where both arguments have the same tracking value, which in both corpora is the most frequent configuration (527/899 tokens in the Besemah corpus, 511/962 tokens in Totoli).

© 2022. John Benjamins Publishing Company
All rights reserved

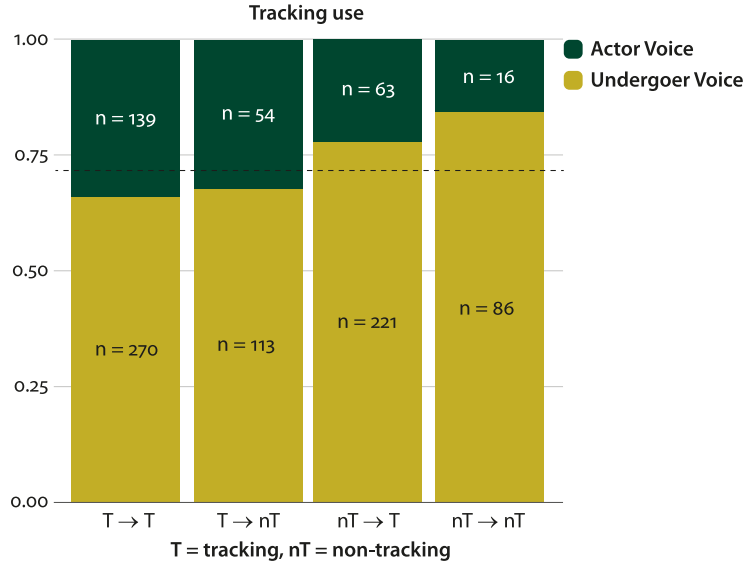

Figure 9. Tracking use and voice in Totoli. Number of transitive clauses = 962

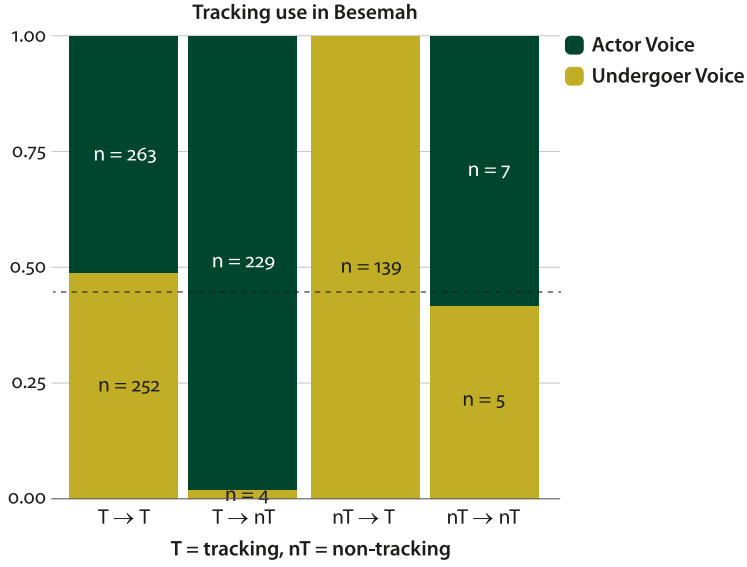

Figure 10. Tracking use and voice in Besemah (McDonnell 2016: 223). Number of transitive clauses = 899

© 2022. John Benjamins Publishing Company
All rights reserved

3.2.4 Generalizability

One further factor that concerns the discourse-referential function of arguments is the GENERALIZABILITY of the referent, that is, whether or not a referent refers to "specific individuals or instances of a category or group" (Ewing 2005: 144). In our corpus, all arguments that received a RefIND/ RefLex annotation were considered to be **particularizing**; all other arguments were counted as **generalizing**. Unlike the TRACKING function discussed in the previous section, which was also determined by the presence or absence of RefIND/RefLex labels, the question whether an argument is particularizing or generalizing is independent of the number of mentions (i.e. **particularizing** includes single referential mentions, while **tracking** requires at least two mentions). Despite this difference, it is, of course, not surprising that the results for these closely related factors are similar. Figure 11 thus shows a very similar pattern to the one shown in Figure 9: it is the status of the actor argument that has a minor impact on voice choice. If the actor is **particularizing**, actor voice is more frequent; if it is **generalizing**, there is a somewhat higher proportion of undergoer voice. As with TRACKING, a generalized linear mixed-effects model analysis revealed a statistically significant association between the GENERALIZABILITY status of the actor (but not of the undergoer) argument and voice (see Supplement for details). The Besemah dataset, like that of Totoli, also shows clear similarities between the GENERALIZABILITY pattern and the TRACKING pattern, as can be seen when comparing Figure 10 and Figure 12 (cp. McDonnell 2016: 225). If one argument is particularizing and the other one is generalizing, then the particularizing one is selected to become the subject. The differences between Totoli and Besemah are again striking and are partially due to differences in coding decisions. As the two factors are highly correlated, we will not further discuss GENERALIZABILITY in the remainder of this paper.

4. Verb/clause-related factors

This section looks at verb- and clause-related factors that might influence voice choice in Totoli. We will investigate GROUNDING and the COLLOSTRUCTION STRENGTH of the verb root.

4.1 Grounding

The basic idea behind the concept of groundedness is that, especially in narratives, some clauses function to advance the main story line and convey information important for the story, while other clauses provide background information

© 2022. John Benjamins Publishing Company
All rights reserved

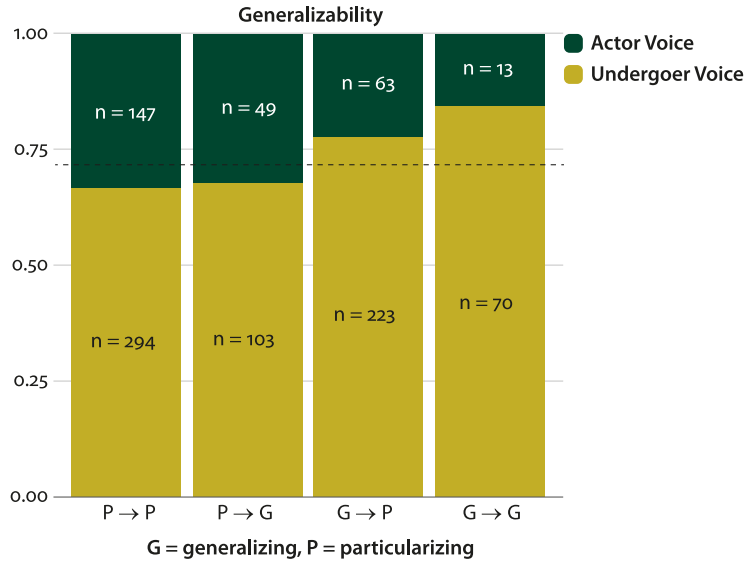

Figure 11. Generalizability and voice choice in Totoli. Number of transitive clauses = 962

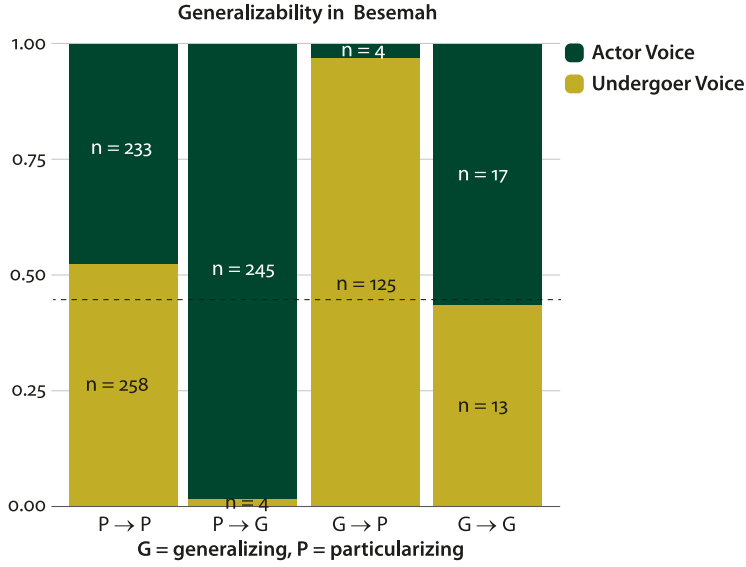

Figure 12. Generalizability and voice choice in Besemah (McDonnell 2016: 225). Number of transitive clauses = 899

or function to set the scene. The main distinction made is thus between clauses that contain foregrounding information and clauses that contain backgrounding

© 2022. John Benjamins Publishing Company
All rights reserved

information. Yet it is not always clear how groundedness can be properly defined, or operationalized. In the literature, two phenomena – DISCOURSE TRANSITIVITY and the use of VALENCY-INCREASING MORPHOLOGY – have been proposed to be useful proxies for determining groundedness. We will discuss each of them in turn.

4.1.1 Discourse transitivity and subordination

One of the early studies that investigates the interaction between groundedness and DISCOURSE TRANSITIVITY on the one hand, and symmetrical voice on the other, is Hopper (1983). Hopper distinguishes three construction types in "early modern" written Malay: the ergative, which corresponds to our undergoer voice, the passive (not considered in this article),[11] and the active, i.e. our actor voice. He hypothesizes that the discourse function of the passive is "a backgrounding one" (Hopper 1983: 71), while the function of the ergative is "to foreground events" (1983: 72). More specifically, it "has a predilection for individuated actions, generally of a concrete, visible kind […], and usually sequenced […]" (1983: 72). The active is considered to be "usual in backgrounded detail when scenic or characterological description is being given" (1983: 79), and is thus, with respect to its discourse function, very similar to the passive.

In order to support his claims regarding the discourse function of the three construction types, he applies the ten transitivity parameters developed and discussed in Hopper & Thompson's **Transitivity Theory** (Hopper & Thomson 1980) to 150 randomly chosen clauses (50 of each construction type).[12] The parameters are: **participants, aspect, kinesis, affectedness of patient, polarity, modality, potency of agent**, **individuation of patient, volitionality** and **punctuality**, and thus include both referent- and event/verb-related features. Note that **agent potency** roughly corresponds to animacy as used in our study ("agents were considered to be potent that is, capable of spontaneous action, if they were human, and not otherwise", Hopper 1983: 75).

According to the basic hypothesis advanced in Hopper & Thompson (1980), backgrounding constructions should have lower transitivity indices than fore-

11. The voice systems found in western Austronesian languages are actually quite diverse. They include both asymmetrical and symmetrical systems, and there is a large number of languages which exhibit both alternation types. These languages have argument demoting voices such as passives and/or antipassives in addition to symmetrical alternations between actor- and undergoer voices (see Kroeger & Riesberg, forthcoming, for an overview on voice systems in Western Malayo-Polynesian and beyond).

12. Hopper's corpus is the *Hikayat Abdullah*, the autobiography of Abdullah 'Munshi' (Abdullah 1932).

© 2022. John Benjamins Publishing Company
All rights reserved

grounding ones. Unsurprisingly, the overall transitivity index calculated for each of the three construction types shows that the ergative construction is the most transitive one (average transitivity count 8.62) and the passive the least transitive (4.78). The active (with an average transitivity count of 5.26) lies between the two, but is clearly closer to the passive than to the ergative (Hopper 1983: 80). Hence, Hopper hypothesizes that undergoer voice (his ergative) is used for foregrounding, specifically for advancing the storyline, while actor voice (his active) is generally used for backgrounded information.

A major problem with this hypothesis is the problem of operationalizing the distinction between foreground and background. Some support for the claim that undergoer voice clauses are more transitive than actor voice clauses in symmetrical voice languages can be found if transitivity is parameterized in the way proposed by Hopper & Thompson (1980), which in itself is not without problems (Lazard 2002, Tsunoda 1985). But in order to correlate transitivity with the foregrounding and backgrounding distinction, one needs a definition of the latter distinction that is independent of the transitivity parameters. As noted by Wouk (1999: 104), a major problem in this regard is that the main storyline is often difficult to identify in non-narrative speech, making it doubtful that it can be applied across different text types.

Pastika (1999) examines foregrounding and backgrounding as a possible factor for voice choice in Balinese, attempting to operationalize the distinction in the following way. He considers all non-main clauses (i.e. complement clauses, adverbial clauses, imperative clauses, interrogative clauses and clauses representing direct speech) as backgrounding (Pastika 1999: 145). Main clauses are considered to be foregrounding if they develop the main story line, which is the primary criterion for foregrounding proposed by Hopper & Thompson (1980: 280f). If this is not the case, these main clauses, too, are classified as backgrounding (Pastika 1999: 145). Pastika also discusses some lexical elements that help to distinguish foregrounding from backgrounding clauses. Thus, he considers certain linking elements, such as *lantas* 'then', *buin* 'again' and *ba/ampun/suba* 'afterwards', as markers for foregrounding material, while he assumes that the use of certain particles signals backgrounding information (Pastika 1999, Section 7.3.2). Pastika shows (1999: 157) that actor voice clauses often report backgrounded information (75% of all actor voices) while foregrounded events are more frequently expressed by undergoer voice clauses (61% of all undergoer voices).

Cumming (1991), like Pastika (1999), finds a correlation between foregrounding – which she calls eventiveness – and undergoer voice in her corpora of Classical Malay and modern Indonesian. In transitive clauses, according to Cumming, eventiveness can be "characterized in terms of high discourse transitivity" (Cumming 1991: 128). In particular, she looks at punctuality, whether the clause

© 2022. John Benjamins Publishing Company
All rights reserved

advances the storyline or not, and sequencing. With respect to the first parameter, Cumming finds that in fact almost all of the clauses in her corpus report punctual events, and that both actor voice and undergoer voice are used (still, in the small number of cases (seven!) where a non-punctual event is described, actor voice is used (Cumming 1991: 129)). Advancing of the storyline is operationalized basically by the distinction between main and subordinated clauses. More than half of the actor voice clauses occur with an adverbial linker, while this rarely happens in undergoer voice clauses (Cumming 1991: 129). Finally, Cumming shows that undergoer voice clauses strongly correlate with the use of the linking element *maka*, which is a marker of "event-line clauses used primarily to indicate sequencing, which may be temporal or logical" (Cumming 1991: 125). She thus concludes that "the functional distinction between [actor voice] and [undergoer voice] in two-argument clauses is based largely on characteristics of the event. […] measures which have to do only with the information status, topicality, or 'topic worthiness' of the actor and undergoer do not serve to distinguish these two clause types" (Cumming 1991: 130).

McDonnell also takes CLAUSE TYPE (main vs. subordinate) as a proxy for groundedness in the Besemah corpus. He distinguishes adverbial and conditional clauses (i.e. all clauses that begin with subordinating conjunctions, such as *ame*/*amu* 'if, when', *antakkah* 'before', *sate* 'after', *sambil* 'while'; McDonnell 2016: 212) from all other clauses, which he considers as main clauses. He finds that less than 10% of all transitive voice constructions in his corpus occur in subordinate clauses (85 clauses in total), which is not surprising given the fact that the corpus consists only of everyday conversational speech. In both clause types, actor voice is more frequent than undergoer voice, but in subordinate clauses the effect is stronger (more than three times as many actor voice as undergoer voice clauses). In addition to CLAUSE TYPE, McDonnell looks at the distribution of actor voice and undergoer voice in declarative, imperative, and interrogative clauses, using sentence mood as another proxy for groundedness. He finds that imperatives have a strong bias for undergoer voice, while interrogative clauses show no clear preference for either of the two voices (McDonnell 2016: 214).

In our study, we distinguished adverbial clauses and main clauses,[13] assuming, as also postulated in the various works quoted above, that adverbial clauses have a backgrounding function in discourse, while main clauses are usually foregrounding. Just like McDonnell's Besemah corpus, the Totoli corpus only exhibits a relatively small number of transitive subordinate clauses (77 instances). However, unlike in Besemah data, there is no significant difference in the distribution of

13. Relative clauses and complement clauses were excluded from our corpus for reasons given in Section 2.

© 2022. John Benjamins Publishing Company
All rights reserved

actor voice and undergoer voice in main and subordinate clauses (cp. Figure 13), as confirmed by a generalized linear mixed-effects model analysis (details in the Supplement).

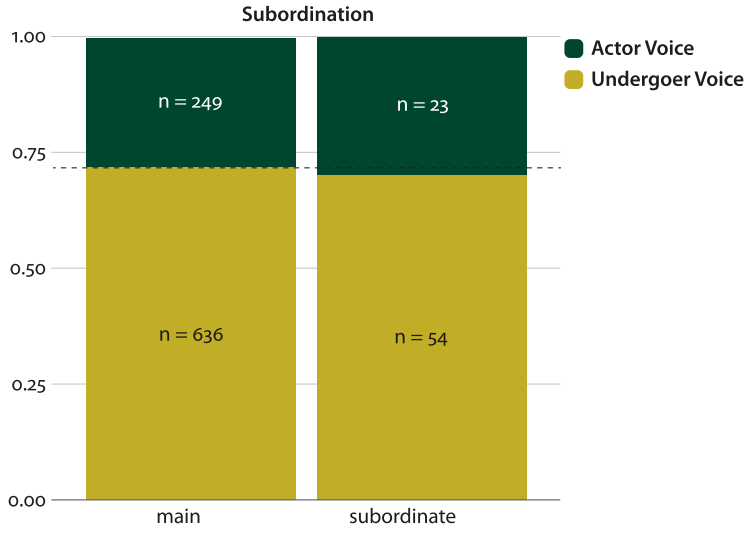

Figure 13. Clause type and voice in Totoli. Number of transitive clauses = 962

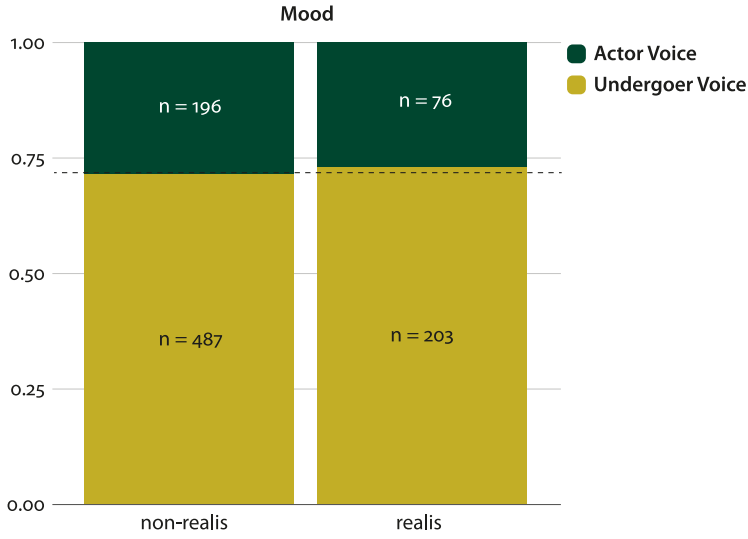

Figure 14. Mood and voice in Totoli. Number of transitive clauses = 962

© 2022. John Benjamins Publishing Company
All rights reserved

We also looked at the distribution of different moods in the two voice constructions (Figure 14). Modality is one of the parameters proposed by Hopper & Thompson (1980) to determine discourse transitivity, who assume that "an action which either did not occur, or which is presented as occurring in a non-real (contingent) world, is obviously less effective than one whose occurrence is actually asserted as corresponding directly with a real event" (Hopper & Thompson 1980: 252). Wouk found that in Jakarta Indonesian non-realis mood in fact tends to occur with actor voice (Wouk 1996: 373).[14] In Totoli, all verbal predicates are obligatorily marked either for non-realis or realis mood (see Section 2.1). As Figure 14 shows, realis and non-realis marked verbs have equal distributions of transitive actor-voice and undergoer-voice clauses in Totoli. Hence this distinction does not seem to be relevant for voice choice. As expected, a mixed effect generalized linear model analysis found no statistically significant association between voice and MOOD.

As a side note, it is worth pointing out that the exact semantic distinctions involved here are still not well understood. As Figure 14 also shows, non-realis mood is considerably more frequent than realis mood, which is one reason why it is considered the unmarked member of the pair. It is thus very likely that we are not dealing with a proper non-realis mood used primarily for future and otherwise imagined events.

While the two potential structural correlates for grounding (CLAUSE TYPE and MOOD) do not turn out to be significant for the Totoli data, we do not believe that this settles the matter. It is very well possible that some of the phenomena included in the somewhat fuzzy notion of grounding are in fact relevant for voice choice. The challenge may consist in properly delimiting these from other, less relevant phenomena also covered by this notion and to provide operationalizable definitions. See Section 8 for further discussion and also the following subsection.

4.1.2 Valency-increasing morphology

Another grammatical feature that has been taken as a proxy for grounding and (discourse) transitivity is the presence of VALENCY-INCREASING MORPHOLOGY. McDonnell (2016) assumes that verbs with transitivizing morphology such as causative or applicative marking are higher in DISCOURSE TRANSITIVITY than verb forms without this kind of marking, and he therefore predicts that Besemah verbs with causative or applicative suffixes are more likely to occur in undergoer voice

14. Note that Jakarta Indonesian does not have a morphological distinction between realis and non-realis mood like Totoli. Wouk coded "conditional, future and negated clauses, imperatives and complements of verbs like *try* and *want* […] as irrealis while all others were coded as realis" (Wouk 1996: 372).

© 2022. John Benjamins Publishing Company
All rights reserved

(see the discussion on DISCOURSE TRANSITIVITY in Section in 4.1.1 above). This hypothesis is borne out in his corpus: in Besemah, both (locative) applicative forms and causative forms are more frequently used in the undergoer voice, while non-suffixed verbs have a stronger tendency for actor voice.

Totoli has two valency-increasing processes: causativization and applicativization. The causative prefix *po-* adds a (new) actor argument to an intransitive or a transitive predicate. Applicativization is slightly more complex. There are two applicative markers which increase the valency of a predicate by one place. Partly depending on the transitivity of the lexical base form, they either add a THEME, a BENEFACTIVE or an INSTRUMENT argument to the verb's argument structure. Both applicative markers occur in all voices and in both realis and non-realis mood. Table 6 summarizes the rather intricate system of applicative formatives in Totoli. There is considerable syncretism between plain voice forms and applicative (voice) marking (cp. Table 1 in Section 2). For a detailed argument supporting the analysis summarized here, the interested reader is referred to Himmelmann & Riesberg (2013).

Table 6. Totoli applicative formatives

	NON-REALIS	REALIS
APPLICATIVE 1 AV	*mo-/moN-/mog-* + *-an*	*no-/noN-/nog-* + *-an*
APPLICATIVE 1 UV (SUBJ = THEME)	*-an*	*ni-* + *-Ø*
APPLICATIVE 1 UV (SUBJ= BEN/INSTR)	*po-/poN-/pog-* + *-an*	*ni-* + *po-/poN-/pog-* + *-Ø*
APPLICATIVE 2 AV	*mo-/moN-/mog-* + *-i*	*no-/noN-/nog-* + *-i*
APPLICATIVE 2 UV	*-i*	*ni-* + *-an*

Figure 15 shows the distribution of the two voices for "plain" voice-marked verbs and verbs with VALENCY-INCREASING MORPHOLOGY. As we can see, undergoer voice is more frequent with causative or applicative verb forms. Figure 16 further differentiates between causatives and applicatives, showing that the latter have a slightly stronger preference for undergoer voice than the former. Yet, numbers for causative formations are low, and the difference we see in Figure 16 between causatives (CAUS) and applicatives (APPL) is not statistically significant. Nevertheless, there is a statistically significant association – though not very strong – between the presence of VALENCY-INCREASING MORPHOLOGY and undergoer voice (see Supplement for details), as in the case of Besemah.

© 2022. John Benjamins Publishing Company
All rights reserved

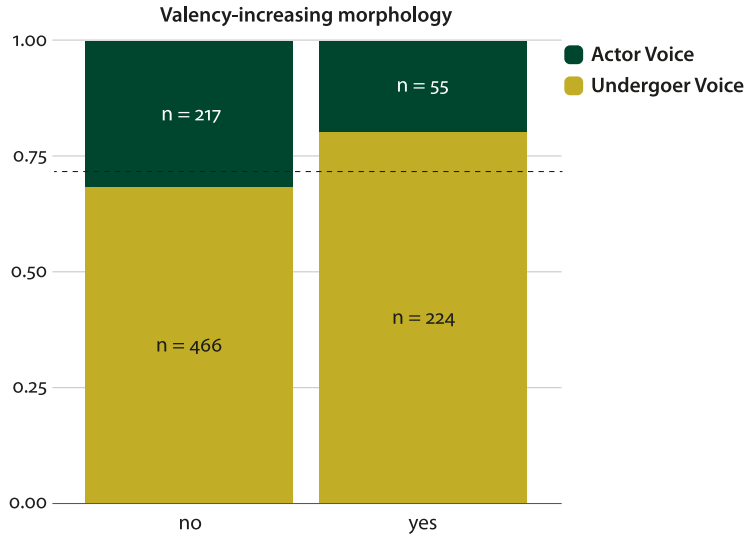

Figure 15. Valency-increasing affixes (causative and applicative) and voice in Totoli. Number of transitive clauses = 962

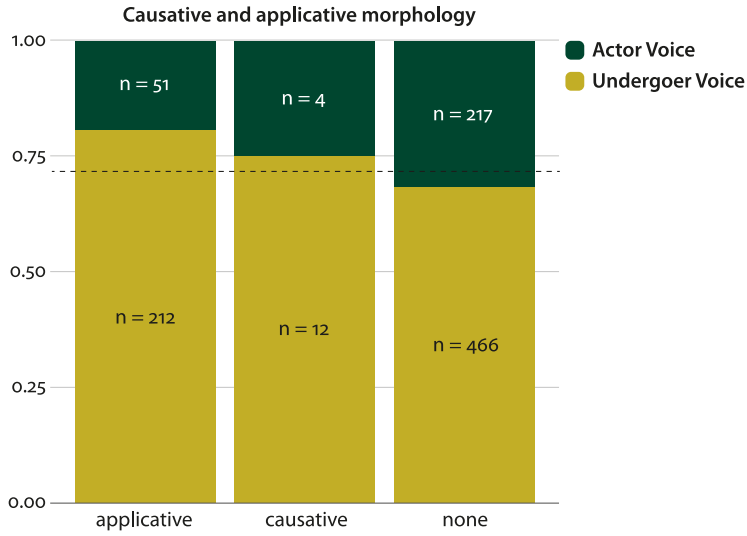

Figure 16. CAUS, APPL and voice in Totoli. Number of transitive clauses = 962

4.2 Collostruction strength

It has been repeatedly observed that verbs in western Austronesian symmetrical voice languages can be biased towards one voice, i.e. not all voices are equally

© 2022. John Benjamins Publishing Company
All rights reserved

well accepted for all verbs. The most extensive study on this topic is probably Latrouite's dissertation on *The coding of prominence and orientation in Tagalog* (Latrouite 2011). Latrouite argues that there are verbs that are inherently oriented towards one of the voices (Latrouite 2011: 114). This inherent orientation does not automatically determine voice choice, but it is a default in those cases where other factors compete or are indecisive. For Tagalog, she argues that all verbs are either actor oriented, undergoer oriented, or neutral in their orientation. This orientation is basically a reflection of the verb's semantic/conceptual structure (cp. Jackendoff 1990, Van Valin & LaPolla 1997, Rappaport Hovav & Levin 1998, among many others). Verbs that exhibit actor orientation are simple (e.g. Tagalog *tawa* 'laugh') and incremental activity verbs (e.g. Tagalog *kain* 'eat'), and verbs of directed motion and manner of motion (e.g. Tagalog *akyat* 'go up', *lipad* 'fly'). Verbs with undergoer orientation are verbs that denote a result with respect to the undergoer argument (e.g. Tagalog *patay* 'kill'), and stative verbs, such as verbs of cognition (e.g. Tagalog *alam* 'know') and perception (e.g. Tagalog *kita* 'see'). Verbs that do not show any inherent orientation towards the actor or the undergoer are verbs of punctual contact (e.g. Tagalog *suntok* 'hit') and transfer verbs (Tagalog *hiram* 'borrow') (Latrouite 2011: 195).

McDonnell also investigates the question of how the internal orientation of a verb root towards either actor voice or undergoer voice is reflected in the overall distribution of voice. He shows that the internal orientation of verb roots is a statistically significant factor for voice choice (McDonnell 2016: 236), but he also admits that it is not easy to interpret these results. That is, **why** a certain verb root is oriented towards a given voice, is not always straightforward to explain (see McDonnell's Section 7.4.3 for discussion).

In our study, we adopted McDonnell's approach to measure the degree to which a verb root is attracted to either the actor voice or the undergoer voice construction by means of a collostructional analysis. For this we used the program Coll.analysis 3.5 (Gries 2014). Our corpus contains 237 different verb roots: Table 7 lists the roots that show the strongest orientation towards actor voice, and Table 8 the ones with the strongest orientation towards undergoer voice.

Looking at the roots listed in these two tables, we can observe tendencies similar to the ones described by Latrouite for Tagalog: The group of verbs oriented towards undergoer voice indeed contains verbs of cognition (*koto* 'know') and perception (*ita* 'see'), as well as verbs that can be argued to focus on the result state of the undergoer (*suang* 'fill', *tau* 'put' and *paas* 'dry').

The group of verbs oriented towards actor voice, on the other hand, consists of simple (*pake* 'use' and *kusut* 'look for') and incremental activity verbs (*gutu* 'make'). The verbs *goot* 'hold' and *salu* 'face' are probably less prototypical activities, but they clearly do not denote states, either (at least, they are not attested

© 2022. John Benjamins Publishing Company
All rights reserved

Table 7. Totoli verb roots with strong actor-voice orientation

Verb root	AV	UV	Collostruction Strength	p
go(l)ot 'hold'	14	3	5.336	< 0.001
gutu 'make'	27	23	4.1262	< 0.001
salu 'face'	7	0	3.8645	< 0.001
pake 'use'	13	7	3.2192	< 0.001
kusut 'look for'	8	2	3.0129	< 0.001

Table 8. Totoli verb roots with strong undergoer-voice orientation

Verb root	AV	UV	Collostruction Strength	p
koto 'know'	0	27	3.9607	< 0.001
ita 'see'	16	96	3.7435	< 0.001
suang 'fill'	0	17	2.4781	< 0.01
tau 'put'	0	15	2.1838	< 0.01
paas 'dry'	0	15	2.1838	< 0.01

with stative morphology in our corpus, and they don't pass any of the language-specific tests for stative predicates, cf. Bardají i Farré et al., submitted). Two further peculiarities need to be mentioned with regard to these two verbs: first, their relatively high frequency is very likely due to the nature of our corpus, in particular due to the stimuli we used to elicit certain kinds of communicative events. All instances of *salu* 'face' in our corpus occur in recordings of the Space Game elicitation game (Levinson et al. 1992), in which speakers were asked to describe pictures that show the orientation of a figure (e.g. a man) towards a ground (e.g. a tree) or towards a certain direction; see (9a) for an example. Similarly, the vast majority of all tokens of *goot* 'hold' (16 of 17 instances) occurs in recordings of the same stimulus task, or in retellings of the Pear Film (Chafe 1980), as exemplified in (9b).

(9) a. *bali tau ampi koig saasalu kita*
bali tau ampii koig RDP2-salu kita
so person side left RDP2-face 2PL
'so the person on the left is facing you'
[spacegames_sequence4_KSR-SP.208]

© 2022. John Benjamins Publishing Company
All rights reserved

b. *mangngana ia googoot alpukat*
-ngo-mangana ia RDP2-goot alpukat
-COLL-child PRX RDP2-hold avocado
'the children are holding avocados' [pearstory_14_SP.049]

Even more important to note, however, is the fact that both *salu* and *goot* frequently do not occur with default voice marking (as introduced in Section 2), but – as also illustrated in the two examples above – in reduplicated forms. These forms clearly exhibit actor voice alignment and have thus been included as actor voice verbs in our statistics. But they certainly cannot be considered prototypical actor voice verbs. They are used to express current states and are most often used in adnominal modifying function (*the child holding the avocados ran away*), rather than as main-clause predicates.

Of the five verbs oriented towards undergoer voice listed in Table 8, three are unambiguously stative verbs (*koto* 'know', *paas* 'dry' and *tau* 'be put'). That is, in their basic use, these verbs denote (intransitive) states, which take an undergoer subject and occur with stative verbal morphology. Example (10) illustrates the use of *tau* 'be put' in its basic function as a stative predicate:

(10) *barang usausatna nattaumo ssaakan dei sia*
barang RDP2-usat=na no-RDP1-tau=mo sasaakan dei isia
goods RDP2-sibling=3SG.GEN **ST.RLS**-RDP1-put=CPL all LOC 3SG
'the goods of his younger sibling are all stored with him' [farming2_0812f][15]

In order to denote a transitive activity and thus be able to take dynamic voice marking, these verbs have to be transitivized, i.e. they have to take applicative morphology. In (11a), the applicative 1 suffix *-an* is added to the stem *tau*, so that it becomes a transitive predicate, taking two core arguments – the agent (not overtly realized here) and the theme *doina* 'his money'. The third participant, the goal argument *kosikang* 'pouch', is realized as an oblique phrase, marked by the locative preposition *dei*. Example (11b) illustrates the ditransitive use of the same verb, in which the applicative 2 suffix *-i* adds the goal argument to the argument structure so that both the theme and the goal are core arguments.

(11) a. *tau-an dei kosikan doi=na*
put-**APPL1**.UV LOC pouch money=3SG.GEN
'(he) put his money into the pouch' [rice2.367]

15. Deviating from the practice in other Totoli examples in this article, Examples (10) and (11b) include more than one intonation unit in a single line.

© 2022. John Benjamins Publishing Company
All rights reserved

b. *ambia ia tau-i niug*
sago PRX put-**APPL2**.UV coconut
'to the sago, coconut is added' [making_ambaa_siote.0704ff]

These verbs, in their basic form, are thus neither actor voice oriented, nor undergoer voice oriented. Rather, they are monovalent stative predicates, and as such they exhibit an undergoer orientation by default, with their subjects always being undergoers. There is no obvious reason why this undergoer orientation is preserved when they are transitivized, but as an empirical fact this is the case. It is relevant here to recall that that verbs with valency-increasing morphology, especially with applicative morphology, are more frequently used in undergoer voice constructions, as demonstrated in Section 4.1.2.

The verb *suang* 'fill' behaves similarly to the above-mentioned verbs *koto* 'know', *tau* 'put, be stored' and *paas* 'dry'. Like these three verbs, *suang*, in its basic use, usually denotes an intransitive state (marked by stative verbal morphology), and is marked by applicative morphology when occurring in transitive or ditransitive contexts (see Examples (12a)–(c)), which are parallel to Examples (10) and (11a) and (b).

(12) a. *mossuangmo i tipi ana kau*
mo-RDP1-suang=mo i tipi ana kau
ST-RDP1-fill=CPL LOC tv MED 2SG
'you are already on TV' [conv_cl.830]

b. *mari=ai kau tiana ku-suang-an dei ngilung=ku*
let=VEN 2SG QUOT 1SG.ACT-fill-**APPL1** LOC nose=1SG.GEN
'let's go, I will just put you in my nose' [story-monkey-python_RSM.056]

c. *Boto' itu ni-suang-an=na minak.*
bottle DIST RLS.UV-fill-**APPL2**=3SG.GEN oil
'He filled that bottle with oil.'

However, and unlike *koto, paas* and *tau, suang* can also be used as a transitive predicate without applicative marking. This is the case in 11 out of the 17 transitive instances that entered our analysis. Its status as a stative predicate, and hence the undergoer orientation of its basic form, is thus not as straightforward as it is for the other three verbs discussed so far. Yet, all verbs in Table 8, including *suang*, conform to Latrouite's hypothesis in that they either denote a result state with respect to the undergoer argument, or are verbs of cognition and perception.

Still, it is not a straightforward task to determine how this finding can be integrated into a comprehensive account of voice choice in symmetrical voice languages. There are two problems. First, as seen in Figure 17 below, the large majority of verbs in our corpus is not clearly actor- or undergoer-voice oriented. If we follow Gries in assuming that only those verbs with a COLLOSTRUCTION

© 2022. John Benjamins Publishing Company
All rights reserved

STRENGTH value larger than 1.30103 exhibit a statistically significant affinity (Gries 2014), then only 22 verb roots in our corpus have a clear orientation towards one of the two voices (14 towards actor voice, 8 towards undergoer voice). All other 215 of the 237 verb roots, almost 91%, show no statistically significant orientation towards either actor or undergoer voice (which in part is simply due to the fact that roots are attested only once or twice). Hence, inherent orientation does not explain voice choice for quite a few verb roots.

The difference in the ratio of oriented and non-oriented verb roots becomes smaller, when looking at actual use in the texts, i.e. when looking at token rather than type frequency. Here we see that the 22 roots that exhibit a clear orientation towards either of the two voices make up 41,3% of all verb tokens in our corpus. While this is quite a considerable amount, it is still the case that the majority verbs used in our corpus are those verbs that show no inherent orientation towards actor or undergoer voice.

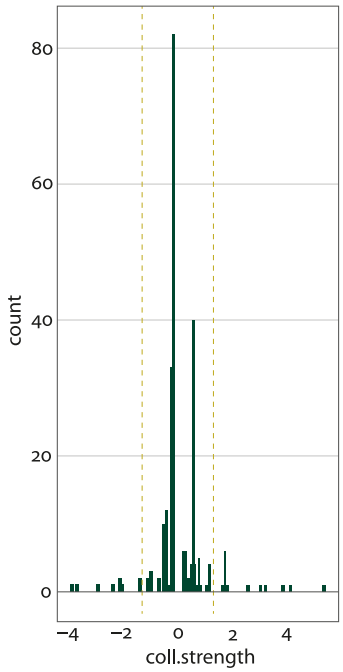

Figure 17. Distribution of verb roots according to collostruction strength

Second, inasmuch as specific lexical items are biased towards one or the other voice, the question naturally arises how this fact can be explained. Simply to say that frequent use in undergoer voice engenders use in undergoer voice is obviously not satisfying, although such a frequency effect certainly contributes to the

© 2022. John Benjamins Publishing Company
All rights reserved

existing biases. Simply to say that some roots are inherently undergoer oriented and others actor oriented is equally unsatisfying, unless such an inherent orientation can be systematically motivated by the event semantics, as proposed by Latrouite (2011). We will briefly return to this issue in Section 8.

In short, the collostructional analysis adds an important piece to the puzzle that until McDonnell's work had been largely overlooked: some verbs are clearly biased towards one or the other voice. Some of these are high frequency verbs (such as *ita* 'see' and *koto* 'know') and thus clearly influence the overall distribution of the two voices (recall that *ita* alone contributes 96 (14%) to the total of 690 uses of the undergoer voice in our corpus). But obviously this does not provide a comprehensive motivation for voice choice in symmetrical languages. Rather, it is part of the explanandum: Why is it that roots show a bias towards one of the two voices?

5. Structural priming

Structural priming describes the tendency for speakers to repeat the syntactic structure that has been used in the previous discourse (see Pickering & Ferreira 2008 for an overview). Since the early 1980s it has been argued that structural priming plays a role in the use of active and passive clauses in English (cp. Weiner & Labov 1983, Estival 1985, Bock 1986), and more recently many studies have shown priming of passive sentences in children (e.g. Bencini & Valian 2008, Branigan & Messenger 2016, Kidd 2012). For western Austronesian symmetrical voice languages, only little is known about the influence of priming on voice choice. In recent studies on the acquisition of the Tagalog voice system, Garcia et al. (2018) and Garcia & Kidd (2020) investigate the effects of structural priming on word order, but as far as we are aware, no experimental research on priming and voice choice is yet available for symmetrical voice languages.

McDonnell (2016: 236ff), following Gries (2005), looks at priming effects in corpus data. In the case of voice choice this means that each instance of a transitive clause is considered a **target**, and the preceding transitive voice-marked construction a prime (cp. McDonnell 2016: 226). We adopted this approach by looking at prime sentences in the preceding five clauses. Targets that constituted the first transitive clause in a text and those where the preceding transitive clause was further away than five clauses were considered to have no prime (cp. Weiner & Labov 1983). As Figure 18 shows, actor voice primes very clearly reduce the frequency of undergoer voices, while undergoer-voice primes trigger a slightly increased use of undergoer voice. When no prime is available, the distribution of actor voice and undergoer voice mirrors the distribution in the overall corpus.

© 2022. John Benjamins Publishing Company
All rights reserved

Note that in general there is a strong tendency for a transitive voice construction to be identical to the preceding one if the latter is no more than five clauses away (out of a total of 633 cases where a prime occurred, 457 involve the same voice).

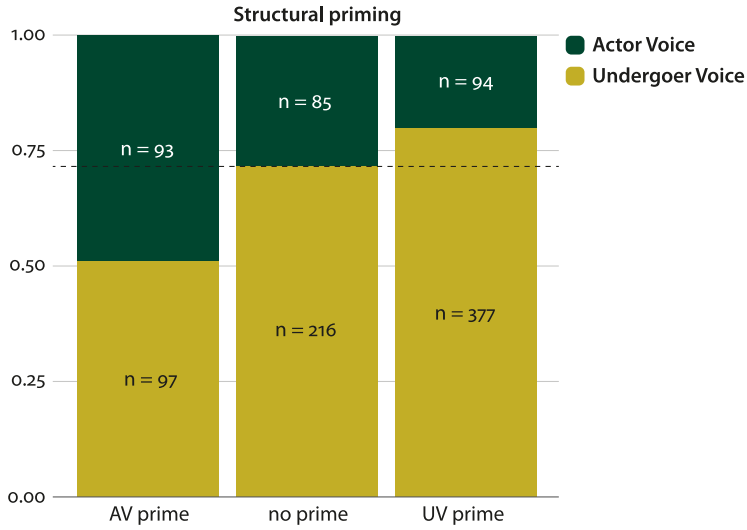

Figure 18. Structural priming and voice in Totoli (no prime = no prime sentence within a distance of 5 clauses). Number of transitive clauses = 962

A likelihood ratio test of a generalized linear mixed-effects model against a null model using the function anova() revealed a statistically significant (overall) association between priming and voice (see Supplement for details).

It is conceivable that other factors influence the priming effect. Therefore, McDonnell (2016) further looks at the prime distance, root identity and speaker identity, predicting that a shorter distance between the prime and the target, the use of the same verb root in prime and target sentences, and the fact that prime and target sentences are produced by the same speaker will strengthen the priming effect. We also investigated these three factors and found that they also influence the priming effect in our corpus. The Supplement provides further details.

6. Interactivity and text type

As far as we are aware, all studies that investigated factors for voice choice in Austronesian languages so far based their research either on corpora solely consisting of narratives (Cooreman 1983, 1988, Cumming 1991, Hopper 1983, Pastika 1999,

© 2022. John Benjamins Publishing Company
All rights reserved

Payne 1994, Quick 2005, Wouk 1996), or on conversational data only (Wouk 1999, McDonnell 2016). Our study is the first of its kind that deliberately draws on a broad range of different INTERACTIVITY and TEXT TYPES, including both monologic and dialogic texts (see Section 2 for details). This allows us to investigate whether INTERACTIVITY and TEXT TYPE influence speakers' use of voice in Totoli.

Our results basically show that neither INTERACTIVITY nor TEXT TYPE appear to have a major impact on voice choice, that is, no statistically significant association between voice and either INTERACTIVITY TYPE or TEXT TYPE was found in the generalized linear mixed-effects model analyses tested for these variables (see Supplement for details). With regard to the somewhat coarse-grained distinction between two basic INTERACTIVITY TYPES, i.e. monologues and dialogues/conversations, no difference in the distribution of voice is discernible in Figure 19 (see Section 2.2 for details regarding the definition of the two types).

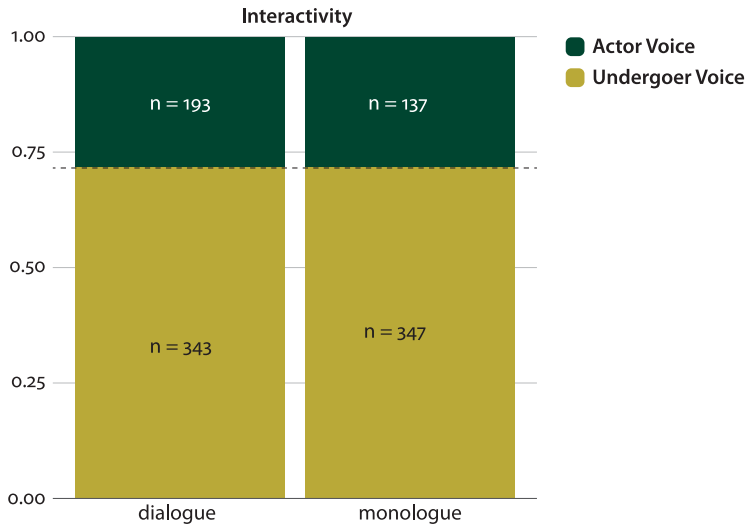

Figure 19. Interactivity (dialogic vs. monologic) and voice in Totoli. Number of transitive clauses = 962

When dividing our corpus into more fine-grained text types, a more nuanced picture appears. As detailed in Section 2.2, we distinguish six text types, defined by their topics and whether or not they involve external prompts: everyday conversations, folk tales, reports of personal experiences, procedural texts, retellings of the Pear Film, and task-oriented interactions (e.g. Space Games). For four of these types the distribution mirrors the overall distribution of the two voices in the corpus (Figure 20). In the case of procedural texts, we find an increased use of undergoer voices. Personal experience reports include an above-average number

© 2022. John Benjamins Publishing Company
All rights reserved

of actor voices. However, the differences just noted are not statistically significant when tested with the same procedure as all the preceding factors.

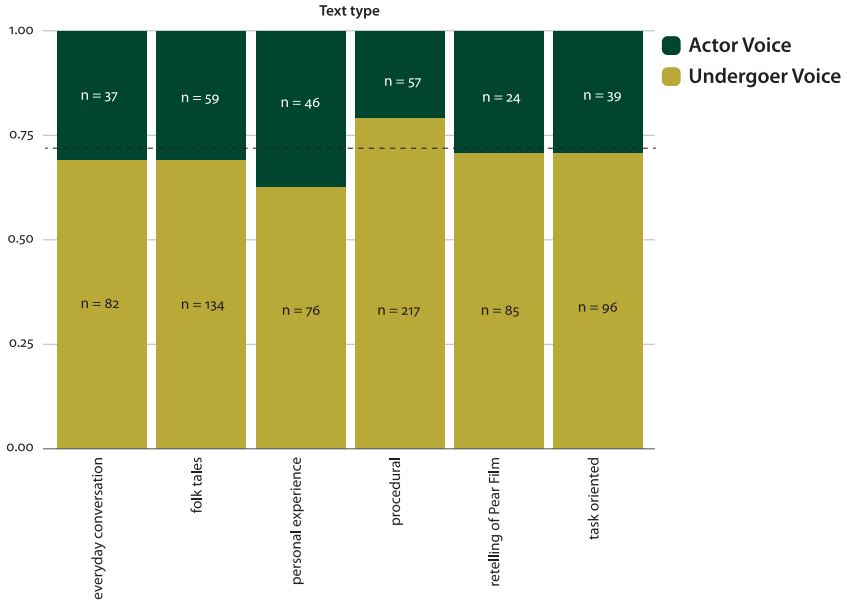

Figure 20. Text type and voice in Totoli. Number of transitive clauses = 962

7. Statistical analysis

In the preceding sections, we made reference to a statistical evaluation of each of the fourteen independent variables we investigated using a generalized linear mixed-effects model.[16] All but five of the fourteen variables were found to be significant according to this statistical test. The exceptions are TOPIC PERSISTENCE, MOOD, CLAUSE TYPE, INTERACTIVITY TYPE and TEXT TYPE.[17]

However, even for the nine variables that were found to be significant it is not clear that they really are major determinants of voice choice in the corpus. In this regard, it should be noted that we generally find the same basic distribution across almost all variables that we investigated. The distributions in the different columns in most figures rarely deviate strongly from the dashed line indicating the

16. We do not include COLLOSTRUCTION STRENGTH as this is itself a statistical measure and, as argued in Section 4.2, is better considered part of the explanadum and not an explanans.

17. As explained in Section 6, in the case of TEXT TYPE a likelihood ratio test was made to assess the overall significance of the variable.

© 2022. John Benjamins Publishing Company
All rights reserved

average 72% to 28% distribution of voices in the corpus. In those cases where more pronounced deviations occur – for example, in the two rightmost columns in Figure 2 ANIMACY – we are dealing with 1% or less of all clauses. In the case of ANIMACY, the configuration **inanimate actor acts on animate undergoer** is attested four times altogether (< 0.5% of the total) and the configuration involving both inanimate actors and undergoers occurs ten times altogether (roughly 1% of the total).

What is perhaps most striking with regard to the distributions observed in our figures is the fact that none of the variables investigated involves a configuration where one voice is used 100% of the time. Even in the case of very small data subsets, such as the ones just mentioned for ANIMACY, we do not find one of the two voices being used in all tokens. In this regard, our Totoli corpus slightly differs from the Besemah corpus, since for the variable TRACKING there is one configuration – **non-tracked actor acts on tracked undergoer** – where only undergoer voice is used. And the converse configuration – **tracked actor acts on non-tracked undergoer** – shows a very strong preference for actor voice.

In our corpus, and in the Besemah corpus as well, distributions generally involve a considerable number of attestations of a minority pattern (20–30%) which cannot easily be qualified as "the odd exception". To stay with the tracking example, an almost equal distribution between actor and undergoer voice is found for the configuration **tracked actor acts on tracked undergoer** in Besemah (Figure 10). In Totoli, one voice is preferred for this configuration, but the other voice is also well attested at a rate in the 30% range (Figure 9). In the Besemah case, it is obvious that tracking status does not provide a motivation for choosing one or the other voice for this configuration, which is the most frequently attested configuration by far (515 out of 899 tokens). But also in the Totoli case, where again it is the most frequent configuration (403/962), the observation that undergoer voice is preferred does not immediately provide a pointer to the basic function of one or the other voice, an issue we will return to in the next section.

In order to further substantiate the impression that the variables investigated here do not predict voice choice well, we decided to further investigate their effect size, expressed in Cramér's V coefficients. The coefficients were obtained through a contingency table analysis using the cramersV() function from the *lsr* package (Version 0.5; Navarro 2015) in R (Version 3.4.0; R Core Team 2017). Table 9 presents a summary of the coefficients for the variables, ordered by effect size in descending order. We exclude those for which no statistically significant association was found in the GLMM test (i.e. TOPIC PERSISTENCE, MOOD, CLAUSE TYPE, INTERACTIVITY TYPE and TEXT TYPE). The data used for the calculation of the coefficients can be found in the figures referred to in the table.

© 2022. John Benjamins Publishing Company
All rights reserved

Table 9. Effect size of variables on voice choice (Cramér's V)

Variable	Matrix size	Cramér's V		Source data
Argument realization	(2×4)	0.28		Figure 8
Priming	(2×3)	0.24		Figure 18
Activation state	(2×4)	0.18	(data subset $n=432$)	Figure 6
Referential distance	(2×3)	0.16	(data subset $n=258$)	Figure 3
Tracking	(2×4)	0.15		Figure 9
Generalizability	(2×4)	0.14		Figure 11
Humanness	(2×4)	0.14		Figure 1
Animacy	(2×4)	0.12		Figure 2
Valency-incr. morph.	(2×2)	0.12		Figure 15

Importantly for current purposes, in all cases for which a statistically significant association was observed, the effect size of the variable can be considered small (following the guidelines for the interpretation of effect size in Sheskin (2011:535) and Cohen (1988:222): $0.1 \leq V < 0.3$ small, $0.3 \leq V < 0.5$, moderate, $V > 0.5$ strong). That is, even in the case of the strongest of the variables, ARGUMENT REALIZATION, the effect size is only in the upper ranges of the SMALL bin. Hence, it would seem fair to say that we have not managed to identify the major factor(s) determining voice choice in Totoli.

There are two options to proceed further from this essentially negative result. On the one hand, it is possible that a combination of some of the variables we looked at provide stronger predictors than any one factor by itself. We will briefly explore this option in the remainder of the current section. On the other hand, it is very well possible that one or more major factors have been overlooked, a possibility we will further discuss in the next section.

As for combining different variables, we first identified existing overlaps between the variables tested so far. In some cases, it is obvious that two variables overlap to a significant degree. A case in point is the overlap between ANIMACY and HUMANNESS, where all humans are animate and most non-humans are inanimate, the difference pertaining to non-human animates. Figure 21 provides an association matrix (showing Cramér's V coefficients for all variables investigated) which allows us to identify overlaps between variables in a more systematic fashion.

The main associations emerging from Figure 21 are unsurprising and were already noted in earlier sections. Next to ANIMACY and HUMANNESS, INTERACTIVITY considerably overlaps with TEXT TYPE. Furthermore, TRACKING overlaps with

© 2022. John Benjamins Publishing Company
All rights reserved

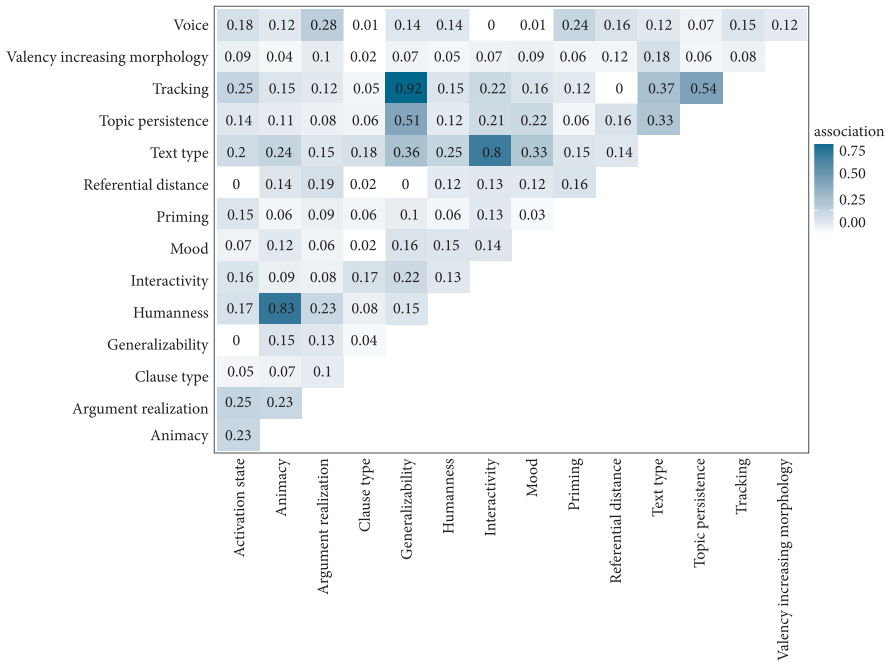

Figure 21. Association matrix (Cramér's V coefficients) of all variables investigated

GENERALIZABILITY and, to a lesser extent, with TOPIC PERSISTENCE. On a lower level, TEXT TYPE overlaps with TRACKING, GENERALIZABILITY, TOPIC PERSISTENCE and MOOD. In order to avoid confounds from strongly overlapping factors, we exclude ANIMACY and GENERALIZABILITY from further consideration. Recall from above that TOPIC PERSISTENCE, MOOD, INTERACTIVITY TYPE and TEXT TYPE are excluded from further consideration for their lack of significance in the GLMM test.

This leaves us with the following seven variables for which it may be productive to investigate the force of combined applications: ACTIVATION STATE, ARGUMENT REALIZATION, HUMANNESS, PRIMING, REFERENTIAL DISTANCE, TRACKING and VALENCY-INCREASING MORPHOLOGY. An in-depth exploration of possible interactions between factors – requiring more sophisticated statistical modelling as well as substantial further scrutiny of the data – is beyond the scope of the present investigation and will be dealt with in a future study. But to provide at least some initial plausibility to further investigations in this direction, we have evaluated the distribution of voice choice under different combinations of five of these

© 2022. John Benjamins Publishing Company
All rights reserved

variables in the decision tree[18] shown in Figure 22 (generated using the function ctree() from the *party* package, Version 1.3.0; Hothorn et al. 2006). ACTIVATION STATE and REFERENTIAL DISTANCE had to be excluded as they are not based on the complete sample of 962 transitive clauses.

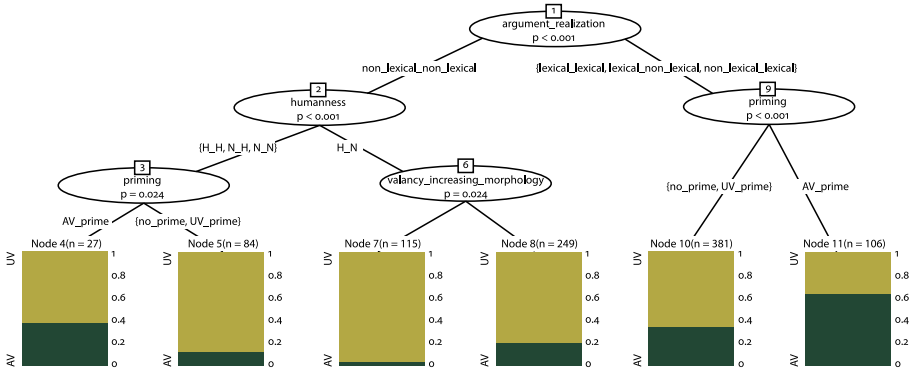

Figure 22. Decision tree of all significant variables that do not strongly overlap and apply to the full data set

The decision tree confirms the result from the Cramér's V test that ARGUMENT REALIZATION is the strongest among all the variables tested. The fact that TRACKING is not included in the result tree indicates that, relative to the other four variables, it hardly plays a role.

Most importantly for current purposes, however, is that the decision tree supports the idea that a combination of variables may allow us to identify contexts where one or the other voice is strongly preferred. This pertains in particular to the third box from the left which shows that undergoer voice is almost always chosen when both arguments are non-lexical (ARGUMENT REALIZATION), the actor is human and the undergoer is non-human (HUMANNESS), and the predicate carries VALENCY-INCREASING MORPHOLOGY. Similarly, the box at the rightmost end deviates strongly from our normal distribution of 72% undergoer voice and 28% actor voice and appears to result from a combination of factors that favour actor voice, i.e. at least one of the two core arguments is lexical (ARGUMENT REALIZATION) and the preceding predicate was also marked for actor voice (PRIMING). Consequently, we expect that a further exploration of combined variables will allow us to delimit

18. Decision trees, also called conditional inference trees, model the outcome variable by binary recursive partitioning. By this, the data is progressively split into groups at specific factor values in order to derive groups with increasingly similar response values. This results in a model that allows to see the effect of the factors and their combinations on the target variable. For an introduction to conditional inference trees, see Strobl et al. (2009).

© 2022. John Benjamins Publishing Company
All rights reserved

more clearly the factor combinations contributing to voice choice in a symmetrical voice language.

8. Discussion

The data and analyses presented in the preceding sections partially confirm and consolidate the results from earlier work on voice choice in symmetrical voice languages, but they also point to the need for a new approach to this issue. They confirm and support the claim that symmetrical voice alternations not only differ structurally from asymmetrical voice alternations (i.e. passives and antipassives), but also differ from them functionally with regard to the discourse conditions of their use. In the following, we will provide more detail and support for this assessment.

In terms of a new approach, two avenues were identified in the preceding section. On the one hand, it appears to be promising to look more closely into the possibility that combined variables (e.g. combining HUMANNESS with VALENCY-INCREASING MORPHOLOGY) allow for stronger predictions of voice choice. This avenue will not be further pursued here. On the other hand, there is the possibility that one or more factors that have not yet been properly identified play a role in voice choice. The main goal of this section is to take a first look at this possibility.

Therefore, we briefly go through each factor group presented in detail in Sections 3 to 6 with the goal of providing an assessment of their potential to make a major contribution to a comprehensive framework for voice choice in symmetrical voice languages. This includes comparisons with voice choice in asymmetrical systems (e.g. active-passive) where voice choice is closely linked to subject choice, as already discussed in the introduction.

We begin with three observations that all works on voice choice in symmetrical systems agree on. First, while not necessarily strictly symmetrical (i.e. 50/50), the usage frequency of the two basic voices in a symmetrical voice system clearly differs from what is found for asymmetrical voice systems. Thus, for example, Biber et al. (1999: 476) report that of all finite verbs in English conversations only 2% are passives. The highest usage of passives (25%) is found in the very specialized register of English academic writing. Compare these figures with the ones given as averages for corpora of symmetrical voice languages in Table 2, which range from 38% undergoer voice and 62% actor voice in Balinese to 72% undergoer voice and 28% actor voice in Totoli.

Second, and more importantly, in the case of English and similar asymmetrical voice systems, the usage conditions for the marked construction (i.e. the passive) can be easily linked to the attested frequencies. Thus, for example, one major

© 2022. John Benjamins Publishing Company
All rights reserved

usage condition for the English long passive, i.e. a passive with an overt actor phrase such as *the jury was approached by many journalists*, is that the undergoer has a higher level of givenness relative to the actor (Biber et al. 1999: 941). The rareness of such passives simply follows from the fact that the configuration **new actor acts on given undergoer** hardly ever occurs in English discourse. In the case of our Totoli data, such a straightforward link between frequency distribution and function is generally not found. This also holds for other studies of symmetrical voice systems.

Third, the universe of discourse in symmetrical voice languages does not fundamentally differ from the universe of discourse in other languages. As just noted, the configuration **new actor acts on given undergoer** rarely occurs in English discourse and thus explains the rareness of the English long passive. This configuration also occurs rarely in our Totoli corpus (and probably in all corpora of natural speech), as seen in Figure 6 where it occurs only six times in a total of 432 tokens for which the ACTIVATION STATE of both core arguments could be determined. Consequently, it is not possible to explain voice choice in symmetrical voice languages by pointing to the fact that the propositional or referential content speakers wish to express is different. In Totoli discourse, actors are more often human, given, tracked, etc. than undergoers, as is the case in other languages. Variation in this regard occurs (e.g. actors are less often made explicit in procedural text types than in other text types), but this again would appear to hold true for most, if not all languages.

Turning now to our brief run through the major factor groups, we already noted in Section 4.2 above that COLLOSTRUCTION STRENGTH is not a determinant of voice in a literal sense. Rather, it appears to be one important aspect of symmetrical voice systems that some verbs are strongly biased towards one or the other voice (but many verbs show only minor biases, if any). A comprehensive model of voice choice has to be able to allow and account for such biases, but strictly speaking, the biases themselves are explananda and do not provide a deeper functional motivation for voice choice.

Roughly the same may be said for VALENCY-INCREASING MORPHOLOGY. There appears to be a link between the transitivity of the state of affairs being expressed and undergoer voice, with increased transitivity correlated with a preference for this voice. The correlation between VALENCY-INCREASING MORPHOLOGY and undergoer voice, which in our corpus is not particularly strong, is one of a number of possible formal manifestations of this link. But it is only a symptom of a potentially systematic link with transitivity. To show that there is indeed a systematic link would require us to be able to show how transitivity and undergoer voice are functionally connected. We will return to this issue below when discussing GROUNDING.

© 2022. John Benjamins Publishing Company
All rights reserved

The role of PRIMING for voice choice is of a very different nature, but it is also on the more shallow side of motivating factors as it is not specific to voice choice. In essence, syntactic priming reflects a universally attested processing bias to reuse linguistic forms, both words and constructions, in ongoing discourse. It is thus not at all specific to voice choice and hence not relevant for explaining the differences in usage conditions between symmetrical and asymmetrical voice systems. Importantly, PRIMING effects may contribute to local clusterings of one or the other voice and, on the lexical level, to COLLOSTRUCTION STRENGTH. However, they are obviously irrelevant whenever no potential prime precedes a particular voice-marked verb.

As for argument-related properties (Section 3), recall from the introduction that voice choice in asymmetrical systems is closely linked to subject choice and argument-related properties are of central importance in this regard. This is not the case in Totoli (and probably most other symmetrical voice languages). There are two lines of argumentation for this assessment, a quantitative and a qualitative one. While the data differ somewhat for the different variables investigated under the heading of argument-related properties (ANIMACY, HUMANNESS, REFERENTIAL DISTANCE, etc.), the basic argumentation can be best illustrated with ARGUMENT REALIZATION. According to the statistical evidence reviewed in the preceding section (Cramér's V, decision tree), ARGUMENT REALIZATION seems to be the one argument-related property that captures best whatever influence argument-related properties may have on voice choice in symmetrical voice languages.

Looking at the raw numbers in Figure 8, repeated here as Figure 23, one could say that with regard to ARGUMENT REALIZATION there is a preference for undergoer voice when both arguments are non-lexical (ca. 85%, right-hand column in Figure 23), a minor preference for undergoer voice when one of the two arguments is lexical (ca. 60%), and a minor preference for actor voice when both arguments are lexical. But what exactly does this tell us with regard to the basic functions of the two voices?

To begin with, we may note that token numbers are unequally distributed across the different configurations (= columns in Figure 23), as is typical for argument-related properties. Roughly half of the tokens (475 out of 962) attest the configuration **two non-lexical arguments**, while less than 5% (43/962) attest the configuration **two lexical arguments**. The second most frequent configuration at roughly 40% is **non-lexical actor acts on lexical undergoer** (third column in Figure 23), while **lexical actor acts on non-lexical undergoer** (second column in Figure 23) is again a somewhat marginal configuration with a little over 5% of the overall total of tokens.

© 2022. John Benjamins Publishing Company
All rights reserved

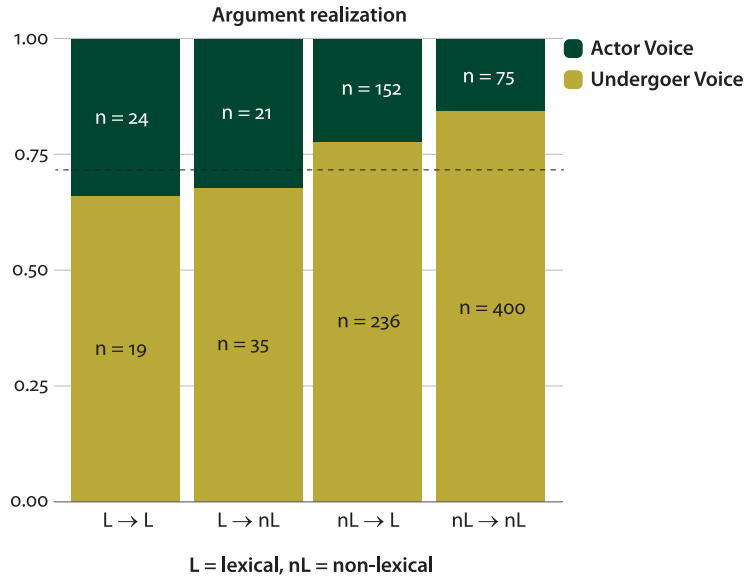

Figure 23. Argument realization and voice in Totoli. Number of transitive clauses = 962

This distribution across the different configurations is not specific to Totoli but can be considered to be typical for spoken language corpora. That is, it is typical for actors to be non-lexically expressed (90% in our Totoli corpus; between 80 and 95% in the 19 corpora discussed by Haig & Schnell 2016: 599). For undergoers, lexical and non-lexical expression is more equally distributed (45% lexical expression in our Totoli corpus; between 40 and 60% in Haig & Schnell 2016: 599). Consequently, it is not the distribution of the two basic referential forms for expressing the macro-roles **actor** and **undergoer** (lexical vs. non-lexical) that is in any way remarkable in Totoli.

What is remarkable from a cross-linguistic point of view is that there is no clear-cut association between either of the two "mixed" configurations (i.e. one argument is lexically expressed, the other one non-lexically; see second and third columns in Figure 23) and one of the two voices. Inasmuch as a non-lexical expression reflects givenness and (discourse) TOPICALITY, and inasmuch as the subject in a transitive event is preferably the more given and topical argument, one would expect a clear preference for one or the other voice depending on which argument is the more given and topical one. That is, the configuration **lexical actor acts on non-lexical undergoer** should show a clear preference for undergoer voice and the converse configuration should preferably be expressed by actor voice. But this is not at all what we find. Instead, we find almost identical

© 2022. John Benjamins Publishing Company
All rights reserved

distributions (60% undergoer voice vs. 40% actor voice) in both configurations. Hence, it is unlikely that voice choice in Totoli is closely linked to subject choice.

Still, there appears to be at least a minor effect of ARGUMENT REALIZATION in that lexical expression of either of the two arguments leads to an increase in actor voice, an effect that becomes stronger when both arguments are lexically expressed. But as this effect occurs regardless of which of the two arguments is lexically expressed, it is also clear that this is again not in any way directly related to subject choice.

The quantitative support for this assessment can be further strengthened by the observation that subject choice does not play a significant role in reference management in Totoli. In European languages, the choice of the subject and the referential expression (lexical, pronoun, zero) makes a major contribution to keeping track of who is doing what in event chains, typically using a lexical expression whenever there is a major switch in agency and/or subject role. This is not the case in Totoli, where instead directional clitics are used to keep track of different referents competing for the actor and subject roles, as Example (13) illustrates.

In this example, which consists of three ((i.)–(iii.)) consecutive clauses of a story, there are two 3rd person singular participants (two boys) who change semantic role and grammatical function in each clause. The whole scene is narrated with boy_2 as the deictic centre. Hence, the giving event in (13i) is marked as involving movement away from the deictic centre (from boy_2 to boy_1), while the giving event in (13ii) involves movement towards the deictic centre (from boy_1 to boy_2). In (13iii), boy_2 moves away from his current position with three pears (described by the speaker as avocados) towards his two friends who are waiting for him further down the road, thereby shifting the deictic centre to this group. Directions toward and away from the deictic centre are marked by the directional clitics *=ko* 'away from deictic centre (andative)' and *=ai* 'towards deictic centre (venitive)'.

(13) i. *i-olatan=**na*** *i-been=**na**=mo=ko* *sapeo*
RLS.UV-wait-APPL2=3SG.GEN RLS.UV-give:APPL1=3SG.GEN=CPL=AND hat
itu
DIST
'he_1 waits for (him_2), and he_2 hands (him_1) the hat'
[pearstory_29_JML.056]

ii. *danna ibeenan**na**mai* *alpukat totolu*
danna i-been-an=**na**=mo=ai alpukat totolu
then RLS.UV-give-APPL2=3SG.GEN=CPL=VEN avocado three
'then he_1 gives (him_2) three avocados' [pearstory_29_JML.057]

© 2022. John Benjamins Publishing Company
All rights reserved

iii. *tissabatuan* *nialamoko* *itu* *danna*
ti-RDP1-sabatu-an ni-ala=mo=ko itu danna
up.to-RDP-one-NR RLS.UV-take:APPL1=CPL=AND DIST then
'one for each (of them he$_2$) brings there' [pearstory_29_JML.058]

The following points in (13) are remarkable from a European reference management and subject/voice-choice point of view. The actor arguments are expressed by pronominal clitics in all instances except in (13iii), where it is zero. The referents in the undergoer role in (13i) and (ii), i.e. the one being waited for (boy$_2$), the recipient of the hat (boy$_1$), and the recipient of the avocados (again boy$_2$), is left unexpressed. The referents referred to by the clitic and 'zero' switch from one verb to the next, which is indicated exclusively by the directional clitics. That is, in this type of example neither choice of subject nor choice of referential expression (zero or clitic pronoun) contribute to reference management. While this does not mean that reference management and subject choice, and as a consequence argument-related properties, do not have any role to play in voice choice in symmetrical voice languages, it clearly shows that these are not pivotal concerns governing usage in such systems.

What is perhaps most remarkable in (13) from a European language point of view is the fact that all verbs in the sequence are in undergoer voice. This is in fact typical for event sequences in narratives in symmetrical voice languages, as documented in greater detail for Tagalog in Himmelmann (1999): whenever there is an extended sequence of actions involving the same major participant(s), it is almost obligatory that these are narrated in undergoer voice. It seems thus very likely that discourse structure has a certain role to play when speakers choose one voice over another.

In the light of this observation, it might come as a surprise that GROUNDING was not among the significant factors in our investigation, even more so as all previous studies that have investigated this factor group (Cumming 1991, Wouk 1996, Pastika 1999 and McDonnell 2016) observe that actor voice clauses are more common in backgrounding contexts, while undergoer voices are more frequently used when describing foregrounding events such as action sequences. This lack of agreement with earlier studies is certainly in part due to the different ways the notion of grounding has been operationalized. In fact, no two studies perfectly agree on their definitions in this regard, partially because the languages differ with regard to potentially relevant formal features that can be coded for. Totoli, for example, is the only language among those that have been studied for voice choice where mood – one of the possible proxies for GROUNDING – is formally marked on verbs. Thus, in this regard, our results are hardly comparable to any of the other studies, where the distinction between foregrounding and back-

© 2022. John Benjamins Publishing Company
All rights reserved

grounding was partly coded manually (i.e. the researcher decided whether or not a sentence is "advancing the storyline"). We believe that this manual approach is highly problematic and runs the risk of involving idiosyncratic and/or biased coding decisions. On the other hand, as mentioned in Section 4.1.1, the exact semantic difference between realis and non-realis marked forms in Totoli is not yet well understood, and thus, though easy to operationalize, using this morphological distinction as a proxy for grounding, too, has to be taken with caution.

Consequently, we believe that the basic hypothesis that discourse structure has a role to play in voice choice is still in need of further scrutiny, even though the results with regard to GROUNDING are problematic for the reasons just mentioned. Besides the action sequences discussed above, the following type of example, which recurs frequently in the corpus, gives rise to this assumption. Here an actor voice construction (*noguadi isia* 'they help him') is followed by two undergoer voice constructions (*suangannamoko* 'they fill (the avocados) for him' and *suangnako* 'they fill them (into the basket)') without there being an easily identifiable motivation for the switch in voice. Note that the switch occurs in a single, fairly complex intonation unit (=fourth line). Note also that a reasonably idiomatic English translation requires that the three boys introduced in the first line continue to function as subjects in all subsequent events.

(14) i. *ana mai totolu moane*
MED come three man
'there are three boys coming along' [pearstory_29_JML.040]

ii. *nog-uad-i isia*
AV.RLS-help-APPL2 3SG
'they help him' [pearstory_29_JML.041]

iii. *ah*
INTJ
'ok' [pearstory_29_JML.042]

iv. *<i><i> nog-uad-i isia suang-an=na=mo=ko <ise>*
AV.RLS-help-APPL2 3SG fill-APPL1.UV=3SG.GEN=CPL=AND
alpukat itu suang=na=ko dei karanjang dei anu dei llenget itu
avocado DIST fill:UV1=3SG.GEN=AND LOC basket LOC FILL LOC basket DIST
'they help him, they fill those avocados for him, they fill (them back) into the *karanjang* the whatchamacallit the basket' [pearstory_29_JML.043]

A closer inspection of such examples shows that they are frequently located at the beginning of episodes, as is the case for Example (14) above. At such a boundary, typically new participants are (re-)introduced, often with a presentative construc-

© 2022. John Benjamins Publishing Company
All rights reserved

tion (existential operator, verb of motion). Then a first event is reported in which the newly introduced participants take part as actors. This is followed by further events where they continue to be actors. Only the first event is in actor voice; the remaining events are in undergoer voice (if transitive).

Based on this observation, we propose the hypothesis that one major context for the use of actor voice is the beginning of a new episode. Undergoer voice, on the other hand, is the preferred choice for event sequences reported within the main body of an episode. A first small pilot study supports this hypothesis. In this study, we took a closer look at the distribution of actor and undergoer voices in the five retellings of the Pear Film in our corpus. Out of 142 transitive clauses, roughly 20% ($n=29$) were in the actor voice and 80% ($n=113$) in the undergoer voice (mirroring quite closely the overall distribution in the total corpus). We then selected one very clear episode boundary that occurred in all five retellings. It describes the scene exemplified in (14) where – after the boy on the bike hit the stone and fell – the three boys appear and help him collecting the pears back into the basket. We found that in the first transitive clauses in these episodes the distribution of the two voices was reversed: 85% actor voice clauses ($n=11$) vs. 15% undergoer voice clauses ($n=2$).

Obviously, the hypothesis that voice choice is related to discourse structure is in need of substantial further scrutiny. There are three major challenges. First, we need to be able to provide operational criteria for identifying discourse-structural units such as episode boundaries without referring to voice marking. This is not an easy task, but we believe that discourse-structural units are more specific and more easily amenable to an operationalizable definition than the vaguer notion of grounding. Second, if there is indeed a statistically significant correlation between voice choice and discourse structure, it is not immediately obvious how this correlation could be explained in functional terms. Using episode boundaries and actor voice again as our example, the challenge here is to go beyond a simple statement of a correlation and provide an explication of the meaning of actor voice constructions in such a way that their use at episode boundaries is motivated by their meaning(s) and structure. Third, in case we actually find a robust correlation between episode structure and voice choice, the biggest challenge will be to identify analogues of episode structure in conversational speech and other genres that do not evince such structure.

9. Conclusion

In this paper we have investigated the question as to which factors determine voice choice in symmetrical voice languages. More specifically, we have asked whether

© 2022. John Benjamins Publishing Company
All rights reserved

voice choice in these languages is linked to subject choice in the same way as it is in asymmetrical voice systems. Here the answer is a clear no. Argument-related properties such as topicality, activation state, animacy, etc. do not play a major role in voice choice in symmetrical voice languages.

We also asked whether the clear structural differences between symmetrical and asymmetrical voice alternations are mirrored in functional differences. Here the answer is a preliminary yes. It seems likely that for symmetrical voice alternations other factor groups are relevant than for asymmetrical voice alternations. However, more work is needed to properly identify and define these other factors groups, one of which may be related to discourse structure (e.g. episode boundaries).

Apart from these main findings, the paper has provided a thorough assessment of all factors that have been proposed to be relevant for voice choice in the literature on symmetrical voice systems. While several of these probably have a minor influence on voice choice, none of them appears to be at the heart of the matter. Importantly, the reasons for why these factors are not highly relevant vary. Priming, for example, is relevant, but not specific to voice choice, reflecting instead a universal processing bias. Collostruction strength differences are part of the explanandum, not the explanans.

Funding

The research for this paper was carried out within the project *B05 Prominence related structures in Austronesian symmetrical voice languages* of the Collaborative Research Centre SFB1252 Prominence in Language (Project-ID 281511265) funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) at the University of Cologne – support we gratefully acknowledge. SR further acknowledges funding from the Australian Research Council as a member of the Centre of Excellence for the Dynamics of Language (2017–2020), and NPH and SR are thankful to the Volkswagen Foundation for long-term funding of their field work in Indonesia. The data in this paper come from the Totoli documentation project funded by the Volkswagen Foundation's DoBeS Program (https://dobes.mpi.nl/projects/totoli/ last access 5 September 2021), and the Totoli documentation corpus collected within the Collaborative Research Centre 1252 Prominence in Language funded by the German Research Foundation (https://lac.uni-koeln.de last access 5 September 2021).

Acknowledgements

Our sincere thank goes to the Totoli community and all the Totoli people who have helped us for many years to learn and understand their language. We are grateful to Maximilian Hörl and T. Mark Ellison for very helpful statistical advice, to Marc Heinrich for the design of the graphs and data annotation, and to Jakob Egetenmeyer for sharing his expertise on episodes and dis-

© 2022. John Benjamins Publishing Company
All rights reserved

course structure with us. Many thanks also to Katherine Walker for data annotation and for very thorough proofreading.

Abbreviations

< >	false start
1	first person
2	second person
3	third person
ACT	actor of an undergoer voice
AND	andative
APPL	applicative
AV	actor voice
COLL	collective
CPL	completive
DIST	distal (demonstrative)
FILL	filler
GEN	genitive
HON	honorific
INTJ	interjection
LOC	locative
MED	medial (demonstrative)
N	nasal
PL	plural
PN	personal name
PRX	proximal (demonstrative)
QUOT	quotative
RDP	reduplication
REL	relative
RLS	realis
SG	singular
ST	stative
UV	undergoer voice
VEN	venitive

References

Abdullah Munshi, bin Abdul Kadir. 1932. *Hikayat Abdullah, Jalid yang pertama* (Malay Literature Series 4). Singapore: Malay Publishing House.

Arka, I. Wayan. 2009. On the zero (voice) prefix and bare verbs in Austronesian languages of Nusa Tenggara, Indonesia, in Bethwyn Evans (ed.), *Discovering history through language: Papers in honour of Malcolm Ross*. Canberra: Pacific Linguistics, 247–70.

Asikin-Garmager, Eli Scott. 2017. Sasak voice. Iowa City: University of Iowa PhD dissertation. Available at https://doi.org/10.17077/etd.dfu5gcvv (last access 5 September 2021).

Bates, Douglas, Martin Maechler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48. https://doi.org/10.18637/jss.v067.i01

Bardají i Farré, Maria, Sonja Riesberg & Nikolaus P. Himmelmann. submitted. Limited-control predicates in western Austronesia: stative, dynamic, or none of the above?

Baumann, Stefan & Arndt Riester. 2012. Referential and lexical givenness: Semantic, prosodic and cognitive aspects. In Gorka Elordieta & Pilar Prieto (eds.), *Prosody and meaning*, 119–162. Berlin: de Gruyter Mouton. https://doi.org/10.1515/9783110261790.119

Bencini, Giulia & Virginia Valian. 2008. Abstract sentence representations in 3-year-olds: Evidence from language production and comprehension. *Journal of Memory and Language* 59. 97–113. https://doi.org/10.1016/j.jml.2007.12.007

Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad & Edward Finegan. 1999. *Longman grammar of spoken and written English*. Harlow: Pearson Education Limited.

Bock, J. Kathryn. 1986. Syntactic persistence in language production. *Cognitive Psychology* 18(3). 355–387. https://doi.org/10.1016/0010-0285(86)90004-6

© 2022. John Benjamins Publishing Company
All rights reserved

Bracks, Christoph A. 2020. The syntax-prosody interface in Totoli. Cologne: University of Cologne PhD dissertation.
Bracks, Christoph A., Datra Hasan, Maria Bardají i Farré, Sumitro Pogi & Nikolaus P. Himmelmann. 2017–2020. *Totoli Documentation Corpus 2*. Language Archive Cologne. Available at: https://lac2.uni-koeln.de/de/ (last access 5 September 2021).
Brainard, Sherri. 1994. Voice and ergativity in Karao. In Talmy Givón (ed.), *Voice and inversion*, 365–402. Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.28.17bra
Branigan, Holly P. & Katherine Messenger. 2016. Consistent and cumulative effects of syntactic experience in children's sentence production: Evidence for error-based implicit learning. *Cognition* 157. 250–256. https://doi.org/10.1016/j.cognition.2016.09.004
Brickell, Timothy C. 2014. A grammatical description of the Tondano (Toundano) language. Melbourne: La Trobe University PhD dissertation. http://hdl.handle.net/1959.9/516057
Chafe, Wallace L. 1976. Givenness, contrastiveness, definiteness, subjects, topics and point of view. In Charles N. Li (ed.), *Subject and topic*, 27–55. New York: Academic Press.
Chafe, Wallace L. 1980. *The pear stories: Cognitive, cultural, and linguistic aspects of narrative production*. Norwood, NJ: Ablex.
Chen, Victoria & Bradley McDonnell. 2019. Western Austronesian voice. *Annual Review of Linguistics* 5(1). 173–195. https://doi.org/10.1146/annurev-linguistics-011718-011731
Cohen, Jacob. 1988. *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.
Cooreman, Ann. 1983. Topic continuity and the voicing system of an ergative language: Chamorro. In Talmy Givón (ed.), *Topic continuity in discourse. A quantitative cross-linguistic study*, 425–489. Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.3.10coo
Cooreman, Ann. 1988. The antipassive in Chamorro: Variations on the theme of transitivity. In Masayoshi Shibatani (ed.), *Passive and voice*, 561–593. Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.16.19coo
Cramér, Harald. 1946. *Mathematical methods of statistics* (PMS-9). Princeton: Princeton University Press.
Cumming, Susanna. 1991. *Functional change: The case of Malay constituent order*. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110864540
Dryer, Matthew S. 1994. The discourse function of the Kutenai inverse. In Talmy Givón (ed.), *Voice and inversion*, 65–99. Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.28.06dry
Estival, Dominique. 1985. Syntactic priming of the passive in English. *Text* 5(1–2). 7–21.
Ewing, Michael C. 2005. *Grammar and inference in conversation: Identifying clause structure in spoken Javanese*. Amsterdam: John Benjamins. https://doi.org/10.1075/sidag.18
Foley, William A. & Robert D. Van Valin. 1984. *Functional syntax and universal grammar*. Cambridge: Cambridge University Press.
Garcia, Rowena, Jeruen E. Dery, Jens Roeser & Barbara Höhle. 2018. Word order preferences of Tagalog-speaking adults and children. *First Language* 38(6). 617–640. https://doi.org/10.1177/0142723718790317
Garcia, Rowena & Evan Kidd. 2020. The acquisition of the Tagalog symmetrical voice system: Evidence from Structural Priming. *Language Learning and Development* 16(4). 399–425. https://doi.org/10.1080/15475441.2020.1814780
Givón, Talmy. 1983. Introduction. In Talmy Givón (ed.), *Topic continuity in discourse. A quantitative cross-linguistic study*, 3–41. Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.3

© 2022. John Benjamins Publishing Company
All rights reserved

Givón, Talmy. 1994. *Voice and inversion*. Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.28
Gries, Stefan Th. 2005. Syntactic priming: A corpus-based approach. *Journal of Psycholinguistic Research* 34(4). 365–399. https://doi.org/10.1007/s10936-005-6139-3
Gries, Stefan Th. 2014. Coll.analysis 3.5. A script for R to compute perform collostructional analyses.
Gundel, Jeanette K., Nancy Hedberg & Ron Zacharski. 1993. Cognitive status and the form of referring expressions in discourse. *Language* 69(2). 274–307. https://doi.org/10.2307/416535
Haig, Geoffrey & Stefan Schnell. 2015. *Annotations using GRAID (Grammatical Relations and Animacy in Discourse)*. Bamberg: Opus. Available at: https://fis.uni-bamberg.de/handle/uniba/21446 (last access 5 September 2021).
Haig, Geoffrey & Stefan Schnell. 2016. The discourse basis of ergativity revisited. *Language* 92. 591–618. https://doi.org/10.1353/lan.2016.0049
Hemmings, Charlotte. 2016. The Kelabit language: Austronesian voice and syntactic typology. London: University of London, SOAS PhD dissertation.
Himmelmann, Nikolaus P. 1999. The lack of zero anaphora and incipient person marking in Tagalog. *Oceanic Linguistics* 38, 231–269. https://doi.org/10.1353/ol.1999.0010
Himmelmann, Nikolaus P. 2001. *Sourcebook on Tomini-Tolitoli Languages. General information and word lists*. Canberra: Pacific Linguistics.
Himmelmann, Nikolaus P. 2005. The Austronesian languages of Asia and Madagascar: Typological overview. In Alexander Adelaar & Nikolaus P. Himmelmann (eds.), *The Austronesian languages of Asia and Madagascar*, 110–181. London: Routledge.
Himmelmann, Nikolaus P. 2010. Language endangerment scenarios: A case study from northern Central Sulawesi. In Margaret Florey (ed.), *Endangered languages of Austronesia*, 45–72. Oxford: Oxford University Press.
Himmelmann, Nikolaus P., Meytal Sandler, Jan Strunk & Volker Unterladstetter. 2018. On the universality of intonational phrases – a crosslinguistic interrater study. *Phonology* 35(2). 207–245. https://doi.org/10.1017/S0952675718000039
Himmelmann, Nikolaus P. & Sonja Riesberg. 2013. Symmetrical voice and applicative alternations: Evidence from Totoli. *Oceanic Linguistics* 52(2). 396–422. https://doi.org/10.1353/ol.2013.0021
Hopper, Paul J. 1983. Ergative, passive, and active in Malay narrative. In Flora Klein-Andreu (ed.), *Discourse perspectives on syntax*, 67–88. New York: Academic Press.
Hopper, Paul J. & Sarah Thompson. 1980. Transitivity in grammar and discourse. *Language* 56(2). 251–299. https://doi.org/10.1353/lan.1980.0017
Hothorn, Torsten, Kurt Hornik & Achim Zeileis. 2006. Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical Statistics* 15(3). 651–674. https://doi.org/10.1198/106186006X133933
Jackendoff, Ray. 1990. *Semantic structures*. Cambridge, MA: MIT Press.
Kidd, Evan. 2012. Individual differences in syntactic priming in language acquisition. *Applied Psycholinguistics* 33(2). 393–418. https://doi.org/10.1017/S0142716411000415
Kroeger, Paul. 1993. *Phrase structure and grammatical relations in Tagalog*. Stanford: Center for the Study of Language.
Kroeger, Paul & Sonja Riesberg. forthcoming. Voice and transitivity. In Alexander Adelaar & Antoinette Schapper (eds.), *The Oxford guide to the Malayo-Polynesian languages of South East Asia*. Oxford: Oxford University Press.

© 2022. John Benjamins Publishing Company
All rights reserved

Latrouite, Anja. 2011. Voice and case in Tagalog: The coding of prominence and orientation. Düsseldorf: Heinrich-Heine-Universität Düsseldorf PhD dissertation.
Lazard, Gilbert. 2002. Transitivity revisited as an example of a more strict approach in typological research. *Folia Linguistica* 36. 141–190. https://doi.org/10.1515/flin.2002.36.3-4.141
Leto, Claudia, Winarno S. Alamudi, Jani Kuhnt-Saptodewo, Sonja Riesberg, Hasan Basri & Nikolaus P. Himmelmann. 2005–2010. *DoBeS Totoli Documentation*. DoBeS Archive MPI Nijmegen. Available at: https://dobes.mpi.nl/ (last access 5 September 2021).
Levinson, Steven C., Penelope Brown, Eve Danzinger, Lourdes De León, John B. Haviland, Eric Pederson & Gunther Senft. 1992. Man and tree & space games. In Steven C. Levinson (ed.), *Space stimuli kit 1.2*, 7–14. Nijmegen: Max Planck Institute for Psycholinguistics.
McDonnell, Bradley. 2016. Symmetrical voice constructions in Besemah: A usage-based approach. Santa Barbara: University of California, Santa Barbara PhD dissertation.
Navarro, Danielle J. 2015. *Learning statistics with R: A tutorial for psychology students and other beginners* (Version 0.5). Adelaide: University of Adelaide.
Pastika, I. Wayan. 1999. Voice selection in Balinese discourse. Canberra: Australian National University PhD dissertation.
Payne, Thomas E. 1994. The pragmatics of voice in a Philippine language. In Talmy Givón (ed.), *Voice and inversion*, 318–64. Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.28.16pay
Pickering, Martin J. & Victor S. Ferreira. 2008. Structural priming: A critical review. *Psychological Bulletin* 134(3). 427–459. https://doi.org/10.1037/0033-2909.134.3.427
Prince, Ellen F. 1981. Toward a taxonomy of given-new information. In Peter Cole (ed.), *Radical pragmatics*, 223–255. New York: Academic Press.
Quick, Phil. 2005. Topic continuity, voice and word order in Pendau. In I. Wayan Arka & Malcolm Ross (eds.), *The many faces of Austronesian voice systems: Some new empirical studies*, 221–242. Canberra: Pacific Linguistics.
R Core Team. 2017. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Available at: https://www.R-project.org/ (last access 5 September 2021).
Rappaport Hovav, Malka & Beth Levin. 1998. Building verb meanings. In Miriam Butt & Wilhelm Geuder (eds.), *The projection of arguments*, 97–143. Stanford, CA: CSLI Publications.
Riesberg, Sonja. 2014. Symmetrical voice and linking in western Austronesian languages. *Pacific Linguistics* 646. Berlin: de Gruyter Mouton. https://doi.org/10.1515/9781614518716
Riester, Arndt & Stefan Baumann. 2017. The RefLex scheme – Annotation guidelines. Vol 14 of SinSpeC. *Working Papers of the SFB 732*. Stuttgart: University of Stuttgart.
Schachter, Paul. 1976. The subject in Philippine languages: Topic, actor, actor-topic, or none of the above. In Charles N. Li (ed.), *Subject and topic*, 491–518. New York: Academic Press.
Schachter, Paul. 1977. Reference-related and role-related properties of subjects. In Peter Cole & Jerrold M. Sadock (eds.), *Grammatical relations*, 279–306. New York: Academic Press.
Schiborr, Nils N., Stefan Schnell & Hanna Thiele. 2018. *RefIND – Referent Indexing in Natural-language Discourse: Annotation guidelines (v1.1)*. Bamberg / Melbourne: University of Bamberg / University of Melbourne.
Sheskin, David J. 2011. *Handbook of parametric and nonparametric statistical procedures*. Boca Raton, FL: Chapman and Hall/CRC Press.

© 2022. John Benjamins Publishing Company
All rights reserved

Strobl, Carolin, James Malley & Gerhard Tutz. 2009. An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological Methods* 14(4). 323–348. https://doi.org/10.1037/a0016973
Svartvik, Jan. 1966. *On voice in the English verb*. The Hague: Mouton.
Tsunoda, Tasaku. 1985. Remarks on transitivity. *Journal of Linguistics* 21. 385–396. https://doi.org/10.1017/S0022226700010318
Van Valin, Robert D. 2005. *Exploring the syntax-semantics interface*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511610578
Van Valin, Robert D. & Randy J. LaPolla. 1997. *Syntax: Structure, meaning and function*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139166799
Weiner, Judith E. & William Labov. 1983. Constraints on the agentless passive. *Journal of Linguistics* 19. 29–58. https://doi.org/10.1017/S0022226700007441
Wouk, Fay. 1996. Voice in Indonesian discourse. *Studies in Language* 20(2). 361–410. https://doi.org/10.1075/sl.20.2.05wou
Wouk, Fay. 1999. Sasak is different: A discourse perspective on voice. *Oceanic Linguistics* 38(1). 91–114. https://doi.org/10.2307/3623394
Zúñiga, Fernando & Seppo Kittilä. 2019. *Grammatical voice*. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781316671399

Address for correspondence

Sonja Riesberg
Laboratoire LACITO
CNRS
UMR 7107 du CNRS
7, rue Guy Môquet (bât. D)
Campus CNRS de Villejuif
94801 Villejuif
France
sonja.riesberg@cnrs.fr

© 2022. John Benjamins Publishing Company
All rights reserved

Co-author information

Maria Bardají i Farré
Universität zu Köln
Institut für Linguistik
Allgemeine Sprachwissenschaft
Köln, Germany
mbardaji@smail.uni-koeln.de

Kurt Malcher
Universität zu Köln
Institut für Linguistik
Allgemeine Sprachwissenschaft
Köln, Germany
kmalcher@smail.uni-koeln.de

Nikolaus P. Himmelmann
Universität zu Köln
Institut für Linguistik
Allgemeine Sprachwissenschaft
Köln, Germany
sprachwissenschaft@uni-koeln.de

Publication history

Date received: 26 August 2020
Date accepted: 18 May 2021
Published online: 27 September 2021

© 2022. John Benjamins Publishing Company
All rights reserved