Dissertation Prospectus: Symmetrical voice constructions in Besemah

Bradley McDonnell

May 13, 2014

1   Introduction

The voice systems of western Austronesian languages have drawn considerable interest from linguists over the past four decades, challenging foundational understanding of various areas of syntactic analysis (e.g., discrete notions of transitivity, the grammatical status of subject and object, etc.). From an Austronesian perspective, the languages of western Indonesia (i.e., Indonesian-type languages after Himmelmann (2002a)) represent an interesting 'transitional' area between the Formosan and Philippine languages (i.e., Philippine-type languages after Himmelmann (2002a)) that have complex voice systems with four or more different voice distinctions (e.g., locative voice, instrumental voice, patient voice, etc.) and the Oceanic languages that in many cases lack voice altogether (Wouk & Ross 2002, Austin 2008).

   In my dissertation, tentatively entitled *Symmetical Voice Constructions in Besemah*, I will provide a thorough treatment of the symmetrical voice system of Besemah, taking a usage-based, constructional approach. For the time being, symmetrical voice is defined as a voice system with multiple transitive voice constructions, none of which is the clear-cut 'basic' form (see Section 2 for more discussion). In the dissertation, I will (i) describe the morphosyntactic properties of grammatical relations and the voice system, (ii) describe the discourse distribution of voice constructions in terms of information structure/information flow and topic continuity based on a corpus of conversational and narrative data, and based on the two previous sections, (iii) provide a statistical model that captures the predictor variables that account for the distribution of voice constructions found in the corpus of Besemah. The primary aim of the study is to clearly describe how speakers of Besemah use symmetrical voice and explain the (probably interacting) grammatical and discourse motivations behind symmetrical voice constructions. A secondary aim of this dissertation is to incorporate complementing functional perspectives: an Austronesian perspective and a typological perspective. There are three major research questions (with several sub-questions) for this study:

1. **What are the structural properties of symmetrical voice constructions in Besemah?**

   (a) What is the status of subject and object?

(b) What is the nature of agentive voice and patientive voice (e.g., active/passive, ergative/absolutive, inverse, or something else altogether)?

(c) What effect does valency increasing morphology (e.g., causatives and applicatives) have on the symmetrical voice constructions?

2. **What are the conditions (discourse and otherwise) in which one voice (e.g., agentive voice or patientive voice) is selected over the other?**

(a) What role do the semantics of the root (if any) play in voice selection?

(b) What role does the syntax of the construction play in voice selection?

(c) What role do discourse properties (e.g., information structure/flow and topic continuity) play in voice selection?

3. **Where does Besemah fit in a typology of voice constructions?**

(a) Where does the symmetrical voice system in Besemah fit in a typology of voice in Indonesian-type languages?

(b) How can the analysis of symmetrical voice in Besemah be extended to other Indonesian-type languages?

(c) How can non-Austronesianists and typologists better understand the place of Indonesian-type languages in terms voice, transitivity, and alignment typologies?

Essentially, these questions (informed by a usage-based, constructional approach) seek to (i) describe the structural properties of symmetrical voice, (ii) explain discourse, cognitive, and social motivations for voice selection, and (iii) inform a typology of voice.

This dissertation contributes to the field of linguistics in a number of ways. First, and possibly most importantly, this dissertation will be the first major grammatical study of Besemah, an under-described language of southwest Sumatra. Not only this, but this dissertation will also be the first major grammatical study of any Malayic language of southern Sumatra, which encompasses dozens of distinct isolects. Utilizing new methods in documentary linguistics, the corpus of naturalistic and elicited data will be archived with The Pacific and Regional Archive for Digital Sources in Endangered Cultures (PARADISEC). Second, the dissertation will contribute to our understanding (or lack thereof) of symmetrical voice constructions. Much of the debate surrounding symmetrical voice constructions revolves around whether these languages show by some (usually non-canonical) definition a nominative-accusative or ergative-absolutive alignment systems. However, this dissertation will not follow these 'categorizing' approaches. Rather, I will follow an approach, whereby I will characterize the constructions in all of their complexity without trying push it into an ergative-absolutive or nominative-accusative box (cf. Croft 2001). Finally, the dissertation will provide a usage-based account of an often overlooked aspect of symmetrical voice constructions, which is the discourse properties of 'speaker choice' in symmetrical voice systems. Using modern quantitative methods (e.g., mixed-effects multivariate statistics), this dissertation will provide a unique perspective on symmetrical voice.

2 Previous research on symmetrical voice

The voice systems (and by extension grammatical relations) of western Austronesian languages have been the center of controversy for over a century (cf. Himmelmann 2005, Foley 2008). Even though this controversy has focused primarily on Tagalog and other Philippine languages (Schachter 1976, Schachter 1977, De Guzman 1992, Kroeger 1993, Naylor 1995, Schachter 1996, Foley 1998), the debate in a number of western Indonesian languages has generated much controversy (Cartier1976, Tchekhoff 1978, Cartier 1979, Hopper 1979a,b, Tchekhoff 1980, Cumming 1987), especially in recent years (Clynes 1995, Wechsler & Arka 1998, Gil 2002, Himmelmann2002a, b, Austin 2001, Arka 2003, Wouk 2004, Arka 2005, Arka 2008). The controversies center around three interrelated themes: `grammatical relations`, `syntactic alignment`, and `voice selection`.

Grammatical relations and syntactic alignment center around two sides of the same issue, because the decision as to which argument (if any) is the grammatical subject and which voice construction (if any) is the 'basic' construction in large part inform the classification of the alignment system. For example, Kroeger (1993) uses several syntactic tests to argue that the NP case marked with *ang* in Tagalog is the subject. Kroeger also implicitly takes the position that there is not a 'basic' transitive voice in Tagalog. Based on this analysis, Kroeger supports a position that Tagalog neither readily fits into nominative-accusative nor ergative-absolutive alignment systems. On the other hand, a number of other scholars do not analyze the *ang* phrase in Tagalog to be the subject, but the absolutive argument (Cena 1979, Payne 1982, Blake 1988, 1993, Liao 2004, Reid & Liao 2004). Instead, these scholars typically follow the pronominal paradigm and analyze certain affixes as applicatives–not voice markers–to show that Tagalog has an ergative-absolutive alignment system. Furthermore, this analysis of Tagalog identifies the agentive (or actor) voice as an antipassive marker, which means that it is not the 'basic' voice construction, at least according to the definition of antipassive in Dixon (1994). For a more detailed discussion of these issues, see Ross (2002), Reid & Liao (2004), and Foley (2008).

The nature of voice in western Austronesian languages has also been investigated at the level of discourse. I refer to investigations of voice at the discourse level as `voice selection` because the issue at hand revolves around the discourse factors that contribute to a speaker selecting one voice construction over another at any given point in a discourse context. Voice selection began to receive some attention in the late 1970's and 1980's with several studies that tried to show how the western Austronesian voice systems were organized at higher levels of discourse (McCune 1979, Cooreman et. al. 1984, Cooreman 1988, Wouk 1989). The majority of these studies have focused on patient prominence in (usually) narrative discourse of western Austronesian languages. There has been a modest increase in the number of studies on voice selection (Brainard 1994 on Karao, Payne (1994) on Cebuano, Wouk (1996) on spoken Jakarta Indonesian, Wouk (1999) on Sasak, Pastika (1999) on Balinese, Huang (2002), and Huang & Tanangkingsing (2011) on Tsou and Seediq). These studies rely primarily on two discourse notions, `topicality` (Givon 1983) and `discourse transitivity` Hopper & Thompson (1980) and usually consider narrative discourse. However, there are some notable exceptions that investigate conversation (Wouk 1989, 1996, 1999). A more recent analysis of voice in colloquial Indonesian by Englebretson (2007) has utilized notions of `stance` to explain patterns of voice selection in conversation. Finally, a number of studies that have

focused on the syntactic properties of voice have taken into account the discourse factors that presumably affect voice selection either as the foundation of their syntactic analysis (Hopper 1979a, b, 1983, 1985, 1988, Rafferty 1982, Verhaar 1983, 1984, 1988) or as supporting their syntactic analysis (Wechsler & Arka 1998, Arka 2003, Arka 2008). However, compared to the vast number of syntactic studies on voice in western Austronesian languages, there are far fewer that focus on voice selection than on the syntactic properties of voice.

2.1   The typology of symmetrical voice in Austronesian languages

Himmelmann (2005) states that '[t]he defining characteristic of these [symmetrical voice] languages is the presence of at least two voice alternations marked on the verb, neither of which is clearly the basic form' (112). The examples from Standard Malay in (1) below demonstrate this property of symmetrical voice.

(1) a. Anak saya me-lihat orang   itu.   [Standard Malay-Indonesian]
         child 1SG   AV-see   person that

         'My child saw that person.'

     b. Orang itu   di-lihat anak saya.
         person that PV-see   child 1SG

         'My child saw that person.'   (Himmelmann 2005)

It is generally thought that the different prefixes, *meN-* in (1a) and *di-* in (1b), mark either the agent, in the prior case, and the patient, in the latter case, as the privileged syntactic argument (i.e., subject or pivot). [1] While it is quite common in a number of the world's languages to be able to treat either the agent or the patient argument as the privileged syntactic argument, there are two other characteristics of the constructions in (1) that make them quite unique. First, each of these constructions is grammatically transitive. That is, both arguments of the agentive and patientive voice constructions appear (at least on the surface) to be core arguments. There is no additional marking that suggests that the patient of the agentive voice (*orang itu* 'that person') or the agent of a patientive voice (*anak saya* 'my child') are oblique arguments. Second, neither construction is identifiably the 'basic' or 'underlying' construction. In both constructions, the verb is marked by a prefix, so it is difficult to posit one construction being derived from the other as in the active-passive voice opposition in English.

It is noteworthy that Himmelmann's defining characteristics of symmetrical voice above (and the definition that I follow here) differs significantly from Foley's original definition of symmetrical voice (Foley 1998). Foley (1998) and Himmelmann (2002a, b, 2005) do agree that the 'lack of an unmarked verbal form, all forms being equally morphologically derived' and 'the CORE status of actor and undergoer arguments regardless of the voice type chosen' are defining characteristics of symmetrical voice (Foley 1998: 73). Foley differs in two

---

[1] The capital letter *N* in the prefix *meN-* represents an underspecified nasal that is homorganic to the first consonant of the root. Generally, if the root begins with voiceless consonant, it is replaced by the homorganic nasal. If the root begins with a voiced consonant, the nasal is homorganic, but does not replace it. See Blust (2004) for further discussion.

regards. First, he states that symmetrical voice has 'the ability of non-subcategorized participants like locatives or instrumentals to freely assume pivot or subject status via their unique voice marking affixes' (Foley 1998: 73). Second, he proposes that there is a dependent relationship between symmetrical voice and precategorical roots in the language. That is, in order to be a symmetrical voice language, a language must have precategorical roots, but the reverse is not necessarily the case. Section 2.2 shows that the first of Foley's latter two defining characteristics unnecessarily excludes Indonesian-type languages, and the second while interesting depends on very controversial and often disagreed upon proposals of precategoricality (Himmelmann 1991, 2008, Gil 1994, Evans & Osada 2005).

Himmelmann's (2005) internal typology of western Austronesian languages (i.e., the non-Oceanic Austronesian languages) highlights the significance that symmetrical voice has in western Austronesia. In constructing his typology, Himmelmann posits two major typological categories, preposed adnominal possession and symmetrical voice.[2] According to Himmelmann, symmetrical voice languages are robust in western Austronesian and include 'the Austronesian languages of Taiwan, the Philippines, Malaysia, Madagascar, western Indonesia (with the exception of Acehnese and the Barrier Island languages) and the northern half of Sulawesi (Saluan (but not Banggai), Kaili-Pamona, Tomini-Tolitoli, Gorontalo-Mongondow, Minahasan and Sangiric)' (Himmelmann 2005: 113).

In earlier typologies of western Austronesian languages, Wolff (1996) and Himmelmann (2002a, b) further classify symmetrical voice languages into Phillippine-type languages and Indonesian-type languages.[3] Philippine-type languages include not only the languages of the Philippines, but extend to the Formosan languages of Taiwan and the languages of northern Sulawesi and northern Borneo in Indonesia. The Philippine-type languages usually display a rich variety of voice alternations, including agent voice, patient voice, locative voice, and circumstantial voice as well as a case marking on NP arguments. Himmelmann (2005) provides a rather strict definition of Philippine-type languages, which states that Philippine-type languages have (i) 'at least two formally and semantically different *undergoer* voices...', (ii) 'at least one non-local phrase marking clitic for nominal expression...', and (iii) 'pronominal second position clitics' (Himmelmann 2005: 113).[4]

Indonesian-type languages include the languages of western Indonesia (Sumatra with the exception Achenese and the Barrier island languages, Kalimantan, Java, Bali, Lombok, and other minor islands in between) usually have a single opposition between agentive (or actor) voice and patientive (or undergoer) voice. According to (Arka 2003), aside from having a symmetrical voice construction, Indonesian-type languages have (i) a true passive construc-

---

[2]`Preposed possessor languages`, where the possessor precedes the possessed within an NP, include non-Oceanic Austronesian languages of Timor, the Moluccas and West Papua as well as the Pidgin-Derived Malay varieties. In this group of languages, there are some that are both a `preposed possessor language` and a `symmetrical voice language` and others that are neither a `preposed possessor language` nor a `symmetrical voice language`.

[3]Himmelmann (2002a, b, 2005) warns that these categories are still in need of further scrutiny. In later work, (Himmelmann (2005) abandons the term Indonesian-type. He gives no explanation for substituting the convenient term Indonesian-type with *languages of western Indonesia*.

[4]According to Himmelmann (2005), his definition 'excludes Malagasy, Chamorro, Palauan and the Austronesian languages of Brunei and Sarawak as well as Tomini-Tolitoli, Gorontalo-Mongondic, Sama-Bajau, and South Mindanao languages, all of which have occasionally been referred to as Philippine-type languages' (113). He also provides a number of other stipulations that are not important for the present discussion.

tion, (ii) an applicative affix that differs from the voice affix that marks voice selection, and (iii) 'voice morphology ... [that] signals linking of generalised roles of A[ctor] or U[ndergoer] to P[ivot]'. While (iii) and to a lesser extend (i) are somewhat controversial, these properties have been generally accepted as defining Indonesian-type languages. However, this definition of Indonesian-type languages and specifically the claims about symmetrical voice should be further scrutinized. The next section looks at the issue of symmetrical voice in Indonesian-type languages more closely.

While there are some properties that cut across Indonesian-type languages, individual languages (and even dialects) differ in the formal properties and the number of contrasts that the voice system distinguishes. For example, Balinese has a two-way distinction between the patientive voice (the bare form) and an agentive voice (marked by a nasal prefix) and Standard Indonesian has two patientive voices (one marked by the prefix *di-* and one that is bare) and one agentive voice (marked by the prefix *meN-*). On the level of discourse, there is still very little known about symmetrical voice in Indonesian-type languages, so it is far too early to draw too many conclusions about the discourse properties of symmetrical voice. The next section discusses the major syntactic properties of voice in a number of Indonesian-type languages, and the subsequent section discusses the discourse properties of voice in several Indonesian-type languages.

2.2 Symmetrical voice in Indonesian-type languages

There have been a wide-range of different analyses of voice in Indonesian-type languages. The majority of these studies have focused on standard varieties of Malay-Indonesian or Balinese. There have been a limited number of studies in other Indonesian-type languages, such as Sasak, Pendau, and Toba Batak to name a few. In what follows, I focus on controversial issues in standard varieties of Malay-Indonesian, essentially whether the transitive clauses show an 'active-passive' or 'ergative' pattern.

Traditional descriptive and/or pedagogical grammars (MacDonald 1967, Dardjowidjojo 1978, Sneddon 1996) and accounts in an older generative grammar framework (i.e., Transformational Grammar) of Standard Indonesian (Chung 1976a, b), for example, have treated the language as if its voice system was similar to English. In these systems, the example in (1a) is the basic active voice, and the example in (1b) is the derived passive voice. In fact, there is an additional patientive voice (only possible for first and second person agents) in (2) that has been referred to by these scholars as the 'passive II' or in Indonesian as the *pasif semu*.

(2) Orang itu   ku=lihat.   [Standard Malay-Indonesian]
     person that 1SG=see

     'I saw that person.'   (Based on Himmelmann (2005))

In this example, the patient NP *orang itu* 'that person' is considered the grammatical subject of the clause (see Chung (1976a, b) for fairly convincing evidence). However, these studies fail to address two foundational issues. First, the core/oblique status of purported non-subject core arguments is often ignored in this literature (i.e., the agent pronoun *ku=* 'I' in (2) and the agent NP *anak saya* 'my child' in (1b)). Essentially, traditional studies assumed these

constructions to be passive without any reference to the demotion of the agent to oblique or the omission of the agent altogether. While it is not possible to demote or omit the agent pronoun in the example in (2), it is possible to demote or omit the agent NP in the (1b). Second, the question as to which construction, agentive or patientive voice, is 'basic' is also ignored; all of these studies assumes that the agentive voice is basic. In order to show that the voice system in Standard Indonesian does in fact show an active-passive alternation, one needs to explain why the agentive voice construction is 'basic' and that the non-subject argument in the patientive voice construction is in fact demoted.

A number of other studies have proposed that varieties of Malay-Indonesian are ergative or have an ergative construction. These studies fall into several categories. First, there are some scholars who have proposed that the bare patientive voice construction exemplified in (2) is ergative (Cartier 1976, 1979, Tchekhoff 1978, 1980) and others who proposed that both forms of the patientive voice are ergative (Hopper 1979a, b, 1983, 1985, 1988, Rafferty 1982, Verhaar 1983, 1984, 1988, Arka 2008). Second, several of these same studies base the analysis of ergativity on syntactic grounds (Cartier 1976, 1979, Tchekhoff 1978, 1980, Arka 2008), while others base their analyses on discourse factors (Hopper 1979a, b, 1983, 1985, 1988, Rafferty 1982, Verhaar 1983, 1984, 1988). A fuller discussion and criticism of the ergative analyses is presented in Cumming & Wouk (1987). For studies prior to 1988, I will repeat many of the important points that are made by Cumming & Wouk (1987).

An example of a purely syntactic analysis that considers the bare patientive voice construction ergative is found in Cartier (1976, 1979) and Tchekhoff (1978, 1980). They both claim that the bare patientive voice in Standard Indonesian is ergative based on a non-standard definition of ergativity originally put forth by Tchekhoff (1978). Tchekhoff's definition of ergativity essentially claims that ergativity is identified by an unmarked modifier (or patient) and unmarked predicate in conjunction with a marked agent. Tchekhoff and Cartier, in turn, take the position that the bare patientive voice predicate and patient subject are unmarked, while the clitic agent is marked.[5] Consequently, this meets Tchekhoff's definition of ergativity. Very similar arguments for ergativity have been proposed for Standard Indonesian (Arka 2008) and Balinese (Wechsler & Arka 1998), usually based on non-standard definitions of ergativity.

Hopper (1983, 1988) exemplify the latter two categories of studies that consider both forms of the patientive voice to be ergative and base their definitions of ergativity primarily on discourse factors. Based on Cartier (1979), Hopper makes a somewhat different claim about ergativity in Early Modern Malay based on discourse notions of transitivity. Hopper proposes that the prefixal patientive voice in Early Modern Malay represents both a passive construction and an ergative construction. Simplifying a bit, the prefixal patientive construction is passive when the patient precedes the predicate with the option of omitting the agent. However, this same prefixal patientive construction is ergative when the patient follows the predicate and the agent is present. Crucial to Hopper's analysis is that the ergative construction serves a foregrounding function, while the passive construction serves a backgrounding function in narrative discourse. In fact, the functional behavior of foregrounding/backgrounding is a more important factor to Hopper than the formal properties

---

[5]According to Cumming & Wouk (1987), she only shows that the clitic agent is marked for first and second singular person forms, but not for third person and not for any plural pronouns (282).

of the prefixal patientive construction.

Essentially, Cumming & Wouk (1987) point out that, in each case, ergativity is *not* defined in reference to both transitive and intransitive clauses (Comrie 1989, Dixon 1994). None of the studies above showed that the single argument of the intransitive clause patterned with patient-like (P) argument of transitive clauses, leaving the agent-like argument (A) in transitive clauses to pattern differently. There are even recent studies that rely on non-standard definitions of ergativity (Arka & Manning 2008). In reference to the discourse-based studies (Hopper 1979a, b, 1983, 1985, 1988, Rafferty 1982, Verhaar 1983, 1984, 1988), Cumming & Wouk (1987) also point out that one cannot always determine whether the clause is ergative or passive, antipassive or active. The status of the clause is rather determined by the discourse status of the clause, the semantics of the clause, or the referential status of the arguments. The distinction between ergative and passive can be quite confusing (see Cumming & Wouk (1987: 283)). The issue here, as Cumming & Wouk point out, is simply that these studies are conflating different levels of linguistic analysis and lack clear morphosyntactic evidence for an ergative analysis.

There have been far fewer studies that investigate the discourse functions of symmetrical voice in Indonesian-type languages. While there is little consensus on why speakers select one voice over the other at any given point in discourse, most studies have found that agentive voice is somehow less transitive than patientive voice. These results are much more clearly demonstrated in narrative discourse (McCune 1979, Cumming 1991, Pastika 1999). Cumming (1991), for example, proposes that patientive voice correlates with `eventiveness`, which she considers to be a component of high transitivity.

In conversational data, this correlation is much fuzzier. For example, Wouk (1989) in a study of spoken (or colloquial) Jakarta Indonesian conversation, finds that agentive voice correlates with lower discourse transitivity and that patientive voice correlates with higher discourse transitivity. She specifically finds that the most relevant factors for discourse transitivity are eventiveness (mood, aspect, foregrounding) and patient status (referentiality, individuation, animacy). Wouk (1989) also found that notions of topicality and thematicity correlated with voice selection.[6] That is, thematic patient arguments commonly trigger patientive voice and thematic agent arguments trigger agentive voice; if both arguments are thematic, then the referent that is more topical triggers the voice.[7] Wouk (1999) looking at both topicality (topic persistence and referential distance) and discourse transitivity finds that Sasak does not show the same patient prominence that is found in other western Austronesian languages (Cooreman et. al. 1984, Cooreman 1988, Payne 1994, Brainard 1994).

Much of the previous research on symmetrical voice in Indonesian-type languages tends to focus on the syntactic properties of voice and rarely makes mention of the discourse properties of voice. These syntactic studies usually rely on constructed examples, running several traditional diagnostics for subjecthood that usually hold for European languages. However, these studies rarely consider examples from naturally occurring discourse. Other studies have

---

[6]Wouk (1989) defines topicality as 'what the section of text is about' and thematicity as 'a referent with a continuous and important role in a section of text' (240).

[7]Wouk (1999: 104), apparently, does not find these correlations in spoken Jakarta Indonesian to be very satisfying for explaining voice selection. Citing Wouk (1989), she states that neither discourse transitivity nor topicality provide striking correlations with different voice constructions.

integrated the discourse and syntax together, but they tend to blur the lines between form and function in such a way that it is often difficult to see whether the language is truly evincing the proposed grammatical pattern (e.g., an ergative-absolutive pattern). Furthermore, these studies do not typically rely on traditional definitions of ergative or active/passive system. In my dissertation, I plan to integrate both the syntactic and discourse properties of symmetrical voice. However, I plan to do so in such a way that clearly distinguishes these different levels of analysis (see Section 4). In the next section, I review the previous research on Besemah and highlight some of the major issues of symmetrical voice in Besemah.

3   Previous research on Besemah

Besemah is an under-described Malayic language, spoken by approximately 400,000 people in southwest Sumatra (Gordon 2005). The language belongs to a complex dialect network that I have recently named South Barisan Malay.[8] Aside from my own work, Besemah has not been the subject of any modern linguistic work. However, during the Dutch colonial period, the Dutch civil servant, O.L. Helfrich compiled proverbs, texts, a short sketch grammar, and a most importantly a dictionary of Besemah and Serawai, a neighboring South Barisan Malay dialect (Helrich 1893, 1895a, b, 1904, 1915, 1921, 1927, 1933). My own research on Besemah has focused on its vowel system (McDonnell 2009), voice system (McDonnell 2010), valency increasing morphology (McDonnell 2013a), clausal nominalization (McDonnell 2013b), grammaticalization of the third person pronoun (McDonnell 2013c), and prosody (McDonnell 2014).

Besemah, like many languages of Southeast Asia, is underspecified for number, case, or gender on nouns, and lacks TAM or agreement morphology on verbs. Tense and aspect are expressed by separate adverbs or particles. Furthermore, arguments of the verb are freely and commonly unexpressed. Word order is basically VO, where the subject can occur preceding or following the predicate (made up of the verb and its object). The genius of Besemah, as it is in probably many if not all western Indonesian languages, is the voice system.

3.1   Symmetrical voice in Besemah

Besemah voice is quite similar to historically conservative varieties of Malay, such as Standard Malay and Standard Indonesian. Some verbal predicates require a verbal prefix as in (3), although some intransitive verbal predicates occur in a bare form as in (4). All nominal predicates are in the bare form, unless they undergo nominalization, as in (5). For clarity, all of the examples in this section are constructed examples. However, in my dissertation, constructed examples will be the exception and not the rule.

(3) a.   Jeme   be-laghi ke ghumah-(ny)e.
          people   MID-run   to house-3

          'People ran to their houses.'

---

[8]These languages have been traditionally referred to as Middle Malay or Central Malay, based on the Dutch term Midden-Maleisch.

b. Die te-tatap   ngah api.
    3   INVOL-touch with   fire

    'He (inadvertently) touched the fire.'

(4) Anak tu   lah lumpat.
     child DET PFV jump

     'The children already jumped.'

(5) Hendi anak.
   H.   child

     'Hendi is a child.'

The prefixes in (3) have several meanings, which I will not go into here. However, the primary meaning of *be-* is as a middle voice prefix in (3a); *te-* in (3b) expresses that the action performed is involuntary or non-volitional. There are also a number of verbs that occur without any prefix as in (4), and the predicate nominal construction has no copula and is simply the juxtaposition of two NPs as in (5). While the focus of this section is on symmetrical voice, the important point here is to note that intransitive predicates do not behave in the same way. This means that, in determining the alignment system, it is not straightforward which construction should be considered the 'basic' intransitive construction. I will consider this question in detail in the dissertation by examining morphological markedness, frequency in discourse, and productivity in discourse. I also leave open the possibility that, based on these criteria, it may not be helpful to identify one of the intransitive constructions as 'basic'.

Symmetrical voice constructions fall into two categories: `agentive voice` and `patientive voice`. Although it remains to be fully fleshed out in Besemah, the idea is that, in agentive voice, the agent argument is the subject and, in the patientive voice, the patient argument is the subject. There is a secondary distinction for the patientive voice between what I term the prefixal patientive voice and the bare patientive voice. As these names suggest, the prefixal patientive voice is prefixed with *di-*, while the bare patientive voice is not marked by any other affix. This classification is exemplified in Figure 1; the formal properties of symmetrical voice with examples are outlined below.

The agentive voice marked by the homorganic nasal prefix *N-* on the verb is exemplified in (6) below.

(6) a. *Jeme*   m-buat   ruti.
         people AV-make bread

         'People made bread.'

     b. M-buat   ruti   *jeme.*
         AV-make bread people

         'People made bread.'

     c. *Jeme*   m-buat-(ny)e.
         people AV-make-3

         'People made it/them.'

Figure 1: Transitive voice constructions in Besemah

```
                        Symmetrical voices
                         /            \
                        /              \
              Agentive voice        Patientive voice
                    |                  /        \
                  (N-)                /          \
                              Prefixal          Bare
                         patientive voice  patientive voice
                                |                 |
                              (di-)            (bare)
```
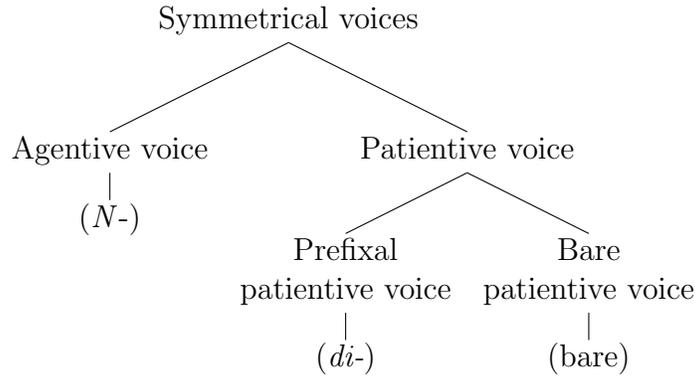
In the agentive voice, the agent argument is the subject and the patient argument is the object. (Arguments for subjecthood and objecthood are discussed below.) The subject *jeme* 'people' may occur before the verb or after the predicate (i.e., the verb and its object). The patient object *ruti* 'bread' and the verb form a 'tight' unit; virtually no adverbs or particles can intervene between them. Additionally, if the patient argument is third person, it may be affixed to the verb as in (6c).[9]

The patientive voice takes on two different forms, the prefixal patientive voice and the bare patientive voice. Both constructions treat the patient argument as the subject and presumably the agent argument as a core argument. (It is unclear, at this point, whether the agent can be considered an object.) The prefixal patientive voice is restricted to third person agents, while the bare patientive voice occurs for every person. The prefixal patientive voice is marked by the prefix *di-* as in the examples in (7). Again, the patient subject can occur preverbally or after the predicate (i.e., the verb and the agent argument). The agent argument is either suffixed to the verb as in (7a), or if it is an NP, then it directly follows the verb as in (7b). The verb and the agent in this construction form a 'tight' unit; nothing can intervene between the two.

(7) a. *Ruti*   di-buat-nye.
         bread PV-make-3

         'Bread was made by him/her/them.'

     b. *Ruti*   di-buat   jeme.
         bread PV-make people

         'Bread was made by people.'

The bare patientive voice is not affixed with a voice marker as in the agentive voice and the prefixal patientive voice. In this sense, it is bare. However, the agent pronoun is either prefixed for the first person singular pronoun as in (8a) or suffixed for the third person pronoun as in (8c). If the agent is first person plural or second person singular or plural, the full pronoun immediately precedes the verb as in (8b); no additional elements may intervene

---

[9]In many western Indonesian languages, person pronouns are commonly considered clitics. For the time being, I consider them affixes. I have not seen compelling evidence for clitichood.

between the agent and the verb. For this reason, a number of scholars have referred to these as agent clitics, despite the fact that they are in no way reduced (Chung 1976a, b). I, again, am not convinced that these are indeed clitics, so I maintain that the reduced forms for first person singular and third person are affixes, and all other pronouns are free forms.

(8) a. Ruti   ku-buat.
         bread 1SG-make

         'Bread was made by me.'

     b. Ruti   kabah buat.
         bread 2SG   make

         'Bread was made by you.'

     c. Ruti   buat-(ny)e.
         bread make-3

         'Bread was made by him/her/them.'

Additionally, the bare patientive voice may also be followed by an NP as in (9) below. In this example the NP directly follows the verb and nothing may intervene between it and the verb.

(9) Ruti   buat jeme.
     bread made people

     'Bread was made by the people.'

Thus far, all of the examples of symmetrical voice have been clearly transitive. However, there are examples of both intransitive agentive voice and patientive voice constructions. The nature of (in)transitivity for agentive and patientive voice differ in significant ways. In order to discuss the intricacies of transitivity in Besemah, I distinguish three types: grammatical transitivity, semantic transitivity, and discourse transitivity. Grammatical transitivity is concerned only with the number of arguments that appear with the verb, whether these arguments appear as NPs or pronominal suffixes on the verb. In naturalistic discourse, this is not always so clear, as arguments of the verb are freely omitted. Semantic transitivity is a prototype view of transitivity, taking into account semantic features related to the agent and patient of the clause (Kittilä 2011). Discourse transitivity is closely related to semantic transitivity, but focuses on how a cluster of semantic features of transitivity is manifested in discourse (Hopper & Thompson 1980, Thompson & Hopper 2001). Features of discourse transitivity are discussed in relation to the dissertation in Section 4.

The intransitive agentive voice constructions are unpredictable in the sense that the root determines the transitivity of the agentive voice construction. Intransitive patientive voice constructions are predictable in the sense that all roots that are able to appear in the patientive voice can either demote or omit the agent if it is third person. First and second person agents in bare patientive voice constructions are less clearly able to be demoted or omit the agent; this is in need of further investigation.[10] The present discussion will be

---

[10]More specifically, I will search more extensively in the corpus and ask speakers if it is possible to have a

concerned mainly with third person forms. Table 1 exemplifies properties of transitivity in Besemah.

Table 1: (In)transitivity of voice in Besemah

	ROOT	AV	PV
LESS TRANSITIVE	*-pulik* 'lie down'	*m-(p)ulik* 'lie down'	——
	-kicik 'say'	*ng-(k)icik* 'say'	——
	-capak 'discard'	*n-capak* 'discard'	——
↕	*-alih 'move'*	*ng-alih 'move'*	——
	-basuh 'wash'	*m-basuh* 'wash'	*di-basuh* 'be washed'
MORE TRANSITIVE	*-buat* 'make'	*m-buat* 'make'	*di-buat* 'be made'
	-tampagh 'hit'	*n-(t)ampagh* 'hit'	*di-tampagh* 'be hit'

Grammatically intransitive roots such as *pulik* 'lie down', *kicik* 'say', and *capak* 'discard' only occur in the agentive voice. It is not possible for these roots to occur in the patientive voice unless they are suffixed with an applicative or causative suffix (e.g., *di-pulik-kah* PV-lie.down-CAUS 'to lie s.t. down' is possible). Grammatically transitive roots such as *basuh* 'wash', *buat* 'make', and *tampagh* 'hit' occur in both agentive and patientive voice without any further affixation. These roots (and any transitive roots that can occur in the patientive voice) can demote the third person agent as in (10a) or omit the third person agent as in (10b).

(10)   a.   Ruti   di-buat   ngah   jeme.
        bread   PV-make   by   people

'Bread was made by people.'

   b.   Ruti   di-buat.
        bread   PV-make

'Bread was made.'

Analogous constructions in related Indonesian-type languages have been referred to as a passive construction. While they have many of the characteristics of a passive construction (i.e., 'promotion' of the patient argument to subject, demotion/omission of agent argument), the discourse function of these constructions is commonly overlooked. For example, why is the agent argument omitted? Is the agent omitted for the same discourse motivations found in passive constructions cross-linguistically? (See Mithun (2008) for these discourse motivations.) Or is it possible that the agent is omitted for the same reason that other arguments of the verb may be omitted?

Grammatical relations and especially the notion of subject in Besemah are crucial to understanding voice in Besemah. However, the notion of 'subject' is not easily defined. The agent of the agentive voice and patient of the patientive voice construction fail many of the traditional diagnostics for subjecthood, including control, raising, relative clause extraction,

---

bare patientive voice form with an agent 'by' phrase that contains a first or second person pronoun. I have examples of this with the third person pronouns.

binding, and extraposition. Some of these categories fail the proposed diagnostic. For example, many varieties of Malay-Indonesian including Besemah do not show clear evidence of control (see Gil (2002) on 'funny' control). Other properties are not appropriate for Besemah. Besemah shows no evidence of true reflexive pronouns; there are words for 'self' that are considered reflexives in other varieties of Malay-Indonesian, but these are not used as such in Besemah. Furthermore, Besemah does not have relative clause constructions with a clear 'gap' (McDonnell (2013c).

The most convincing piece of evidence for a privileged syntactic argument is the nature of 'quantifier float'. 'Quantifier float' has been used by Schachter (1976, 1977) and Kroeger (1993) for Tagalog as a diagnostic for subject arguments, while Arka (2008) and Arka & Simpson (2008) for Balinese, and Musgrave (2000, 2001) for Standard Indonesian have used it to test for core arguments (or term arguments in the terminology of Lexical-Functional Grammar). 'Quantifier float' provides the clearest criterion for subjecthood in Besemah. In this sense, Besemah is quite rare among the Indonesian-type languages, as I am not aware of any other Malayic language that restricts 'quantifier float' to subjects. To begin, consider the following examples of the quantifier *gale* 'all' in (11)–(14). In these examples, the argument that the quantifier targets is in bold face.

(11) Matik gale **die tu**.
      die   all 3   that

      'All of them died.'

(12) Bilang **jeme** lah ng-(k)icik-kah diwik gale,
      every person PFV AV-say-CAUS   self all

      'Everyone already gossiped about her,'

(13) P: Be-cakagh=lah   singe-nye **lanang** banyak tu   singe-nye.
         MID-search=EMPH   result-3   male   many   that result

         'There were a lot looking as a result there were lots of men.'

      H: Di-kinak-i   gale ai.
         PV-see-APPL   all ah

         'All (of the men) were looked at, ah.'

(14) J: Die ni   anu galak ng-ambik-i-nye **kiung**.
         3   this uhm want AV-take-APPL-3 snail

         'He uhm wanted to take them, the snails.'

      D: Adak di siring masih ku-ambik gale di siring.
         but at river still 1SG=take all at river

         'The ones in the river I'll still take them all.'

The intransitive verb *matik* 'die' in (11) shows the quantifier immediately following the predicate, while the target of the quantifier and the single argument *die tu* 'they' follows both the predicate and the quantifier. As noted earlier in this section, the subject NP *die tu* 'they' could also occur before the predicate. In this case, it would still be the target of the quantifier. Most importantly, the subject here, despite its placement, is the target for the quantifier. In (12), in the agentive voice construction, the quantifier occurs at the end of the clause, targeting the semantically agent subject *jeme* 'people' and not the patient object *diwik* 'her'. In fact, in transitive clauses this is by far the most preferred position of the quantifier. In the case of the prefixal patientive voice in (13) and the bare patientive voice in (14), the quantifier is post-predicate and the patient subject argument in each example is quantified, *lanang* 'men' in (13) and *kiung* 'snails' in (14). Notice that in both cases, the patient subject is actually not found overtly in the same clause as the quantifier. Rather, the targeted NP is found in the previous utterance of another speaker.

The grammatical relation of object is far less clear. There have been some attempts to define objects in Standard Indonesian in Kana (1986) and Chung (1976a), but Musgrave (2001) points out that these are problematic or inconclusive at best. The only diagnostic that Musgrave (2001) finds reliable is 'quantifier float', which only applies to subjects in Besemah. The bigger question here may not be whether or not Besemah has objects: it may be more related to the core/oblique status of non-subject arguments (i.e., is the agent argument in patientive voice construction core or oblique?)

This overview of Besemah voice outlines the basic properties of symmetrical voice in Besemah and highlights many of the unresolved issues surrounding symmetrical voice in Indonesian-type languages. In my dissertation, I plan to address many of these unresolved issues to get at the larger questions surrounding symmetrical voice (e.g., What is symmetrical voice? Is symmetrical voice a useful way to look at Indonesian-type languages? Is it better to view these languages in other terms, such as ergative(-absolutive) or (nominative-)accusative?). However, in concluding this section, it may be necessary to qualify the term symmetrical voice. Symmetrical voice in Besemah and probably most other western Austronesian languages is not strictly symmetrical. Arguments of the verb within a single construction do not behave in exactly the same manner, and most crucially here arguments of the verb across transitive constructions do not behave in exactly the same way (i.e., agentive and patientive voice constructions are not exact mirror images of each other). Symmetrical voice rather refers to a more approximate similarity, rather than some sort of strict identity. The question, then becomes, at what point is a language no longer considered to possess a symmetrical voice system. In this dissertation, I plan to better assess this question with evidence from Besemah.

4 Overview of dissertation

While a number of studies of symmetrical voice have focused either on a purely syntactic analysis or a discourse-based analysis that often lacks a solid syntactic basis (i.e., discourse ergativity), this section will lay out how this dissertation aims to provide clear syntactic argumentation for the status of symmetrical voice and clear discourse motivations for the different symmetrical voice constructions.

4.1   Part I: Introduction

Chapter 1: Symmetrical voice: Problems and prospects   The first chapter introduces the nature of symmetrical voice in the languages of western Austronesia, focusing on Indonesian-type languages. I will present the previous literature on symmetrical voice and problematize the previous analyses that were presented in Section 2. I will focus on definitions of symmetrical voice (Foley 1998, Himmelmann 2002a, b, 2005) and the defining characteristics of symmetrical voice languages, presenting examples from Standard Indonesian, Balinese, Sasak, varieties of Javanese among other languages.

Chapter 2: Usage-based and constructional approach   The second chapter of the dissertation will lay the theoretical foundations of the dissertation. That is, it will outline the usage-based (or discourse-functional) approach to investigating language structure (Bybee 2010) as well as a constructional approaches to syntax (Goldberg 2006). There are a number of advantages to a constructional approach (i.e., strongly grounded in cognition, easily integrated with usage-based approaches, looks at the entire grammar of the language, etc.). One example of its usefulness for this study is the inheritance hierarchy. A constructional approach recognizes that, on the one hand, constructions represent a number of polysemous meanings and that no two constructions are synonymous, and, on the other hand, constructions are cognitively related in an ordered manner. These complementary principles can help account for the facts concerning voice in Indonesian-type languages. It has often been noted that the agentive voice prefix in Malayic languages occurs in both intransitive and transitive clauses. Some studies have tried to unite these seemingly different functions under a single abstract function, which in the end fails to recognize sizable differences between the two constructions. Others who treat these as two completely different constructions do not recognize the obvious formal and functional similarities. The same could be said for the prefixal patientive voice. Arka & Manning (2008) have argued that without the presence of the agent enclitic, the prefixal patientive voice construction is a passive construction, but with the presence of an agent enclitic, it is an ergative construction. While it may turn out that there is some truth to this statement, there is still a clearly general function (i.e., both marking the patient as the subject) that unites these functions. Constructional approaches recognize that it is not all or nothing; constructions that show differences are polysemous, but these same constructions may be ordered in a cognitively coherent manner.

Chapter 3: Grammar sketch of Besemah   This chapter lays out the basics of Besemah morphology and syntax. This chapter will primarily treat issues that are crucial for the analysis or robust categories in the language. The section will include a description of basic phonology (phoneme inventories and morphophonemics), lexical categories (at the root- and word-level), basic derivational morphology, basic clause structure, noun phrase structure, topicalization, clausal nominalization, predicate nominals among other topics. The grammar sketch will closely follow those found in Adelaar & Himmelmann (2005).

4.2   Part II: The morphosyntax of symmetrical voice

Chapter 3: Grammatical relations and symmetrical voice   This chapter presents the criteria for subjecthood and objecthood in Besemah. As section 3.1 shows, these grammatical categories in Besemah are not at all straightforward. I will present counter-evidence to many of the commonly assumed properties of subjecthood (e.g., control, raising, relative clause extraction, binding, and extraposition), which has been widely used in other studies of western Indonesian languages (Wechsler & Arka 1998, Arka 2008). Then, I will present some evidence for the category of subject from 'quantifier float'. I will also present a more limited discussion of objects in Besemah, following Musgrave (2001). As mentioned in section 3.1, the discussion may be more concerned with non-subject core arguments than with objects. In this section, I will rely primarily on syntactic analysis of naturally occurring discourse. I think it is important to show that such phenomena actually occur in discourse. I am, however, not opposed to including some targeted elicitation either based on examples of naturalistic discourse or picture elicitation tasks. These examples will be in the minority in the dissertation.

Chapter 4: Transitivity and symmetrical voice   This chapter presents the complexities of transitivity for symmetrical voice. In this chapter, as I did in section 3.1, I will distinguish three types of transitivity: grammatical transitivity, semantic transitivity, and discourse transitivity. In particular, I will investigate the extent to which agentive and patientive voices are variably transitive in discourse. I will be especially interested in the purported non-subject core arguments (i.e., agent in patientive voice construction and patient in agent voice construction). Do these arguments show properties of referents in highly transitive constructions (e.g., definite, highly individuated, highly affected)? This is crucial to understand if symmetrical voice constructions are really symmetrical in discourse or if symmetrical voice constructions are only made to seem symmetrical in elicitation. An important component of transitivity will also look at some of the root semantics and combinatorial possibilities of symmetrical voice constructions. If time and space permits, I would also like to discuss valency increasing morphology in this same section (i.e., causative and applicative suffixes).

4.3   Part III: Voice selection in discourse

Chapter 5: The information structure/flow properties of voice   This chapter presents an analysis of the informational structure/flow properties of agentive and patientive voice in Besemah. This chapter will be largely qualitative, explaining the variables that are coded in the corpus of Besemah (see below). These properties include information structure and information flow properties, such as activation state, identifiability, identifiability pathway, generality (Chafe 1994, Lambrecht 1994, Du Bois & Thompson 1991, Ewing 2005). These properties may be complemented by the discussion of discourse transitivity in the previous chapter.

Chapter 6: The statistical model of voice   This chapter will take many of the different semantic, syntactic, and discourse variables discussed above to perform a quantitative

analysis of the Besemah corpus. The corpus of Besemah currently consists of approximately 5 hours of transcribed, translated, and glossed conversational data. There is an additional hour of narrative data. I hope to expand both the annotation of the current corpus and add additional conversational data to the corpus during several field trips from May 2014 to April 2015.

The data will be analyzed using a generalized mixed effects binary logistic regression. The independent variable will be VOICE with the levels *agentive voice* and *patientive voice*. As I showed in Section 2, these voices can be quite variable, so one may question the use of a binary distinction between agentive and patientive voice. While I will code the corpus for all of the distinctions discussed above, I will begin with two macrocategories. If the dependent variable needs to be split up, I will have the coding sufficient to do so. Note that I will only code examples of symmetrical voice in the corpus and not every clause or IU.

There will be a number of possible independent variables, which are listed in Table 2–6. The independent variables are taken from formal properties of symmetrical voice constructions (Table 2), verbal semantics (Table 3), information flow (Table 4), discourse transitivity (Table 5), and topic continuity (Table 6). These lists represent far more independent variables than will be necessary or even possible for this study. I will begin coding the data for as many of these variables as possible. As I code, I am expecting that a number of variables will turn out to be too difficult to operationalize or become clearly irrelevant as a predictor for voice selection. At the same time, I am open to replacing some of these variables with other, more helpful variables. Finally, some of these factors will not be included in the statistical model, but will be factors that I code for in order to get a better handle on the qualitative analysis. For example, WORD ORDER, as it is operationalized below, will not make a good independent variable because there are too many levels that will occur very infrequently. However, I would like to keep track of word order and may be able to base another more useful variable on it (e.g. VERB INITIAL with the levels *yes* and *no*).

4.4   Part IV: Implications

This chapter will largely depend on the analysis of the previous chapters. Therefore, these last two chapters (or possibly a single chapter) are tentative.

Implications for Austronesian linguistics   This chapter will lay out the implications of the study for Austronesian linguistics and primarily Indonesian-type languages. More specifically, I hope to provide a better understanding of symmetrical voice systems. What are some defining characteristics? Are there any truly symmetrical voice languages? What makes Indonesian-type languages different?

Implications for typology   The final chapter will lay out the implications of symmetrical voice for typology. What does symmetrical voice tell us about alignment systems and voice systems, if anything at all? Where does symmetrical voice fit in the typology of voice systems? How can typologists better understand symmetrical voice?

5 Research plan

The research plan is as follows. From May 2014 to April 2015, I will be on fieldwork in Indonesia on a Fulbright-Hays Doctoral Dissertation Research Abroad scholarship. I will be based out of the Center for Languages and Cultures at Atma Jaya University and the Max Planck Institute for Evolutionary Anthropology's Jakarta Field Station. From there, I plan to take three field trips to the Besemah village of Karang Tanding. During these field trips, I will record additional conversations in addition to transcribing, translating, glossing, and coding, the conversations. Upon returning in Spring 2015, I will prepare the data for the statistical analysis and begin writing the chapters on grammatical relations and symmetrical voice and transitivity. During the summer of 2015, I plan to present some of my dissertation at the International Conference on Austronesian Linguistics in Taipei and work on the grammar sketch. In Fall 2015, I plan to work on the chapters on information flow as well as the statistical analysis of voice selection. In Winter 2016, I plan to work on the introductory chapters and hopefully present my work at the LSA. In Spring 2016, I plan to write the concluding chapters and finish any revisions. I also plan to present my dissertation colloquium talk in Spring 2016.

References

Adelaar, K Alexander & Nikolaus Himmelmann, eds. (2005) *Austronesian Languages of Asia and Madagascar*, *Routledge Family Language Series*, vol. 7. New York, NY: Routledge.

Arka, I Wayan (2003a) *Balinese morphosyntax: A lexical-functional approach*, vol. 547. Canberra, Australia: Pacific Linguistics.

Arka, I Wayan (2003b) Voice systems in the Austronesian languages of Nusantara: Typology, symmetricality and Undergoer orientation. *Linguistik Indonesia* **21**(1): 113–139.

Arka, I Wayan & Christopher D. Manning (2008) Voice and Grammatical Relations in Indonesian: A New Perspective. In *Voice and grammatical relations in Austronesian languages*, Simon Musgrave & Peter K. Austin, eds., Stanford: Center for the Study of Language (CSLI), 45–69.

Arka, I Wayan & Malcolm Ross (2005) *The many faces of Austronesian voice systems: some new empirical studies*. Pacific Linguistics, Research School of Pacific and Asian Studies, ANU.

Arka, I Wayan & Jane Simpson (2008) Objective voice and control into subject clauses in Balinese. In *Voice and grammatical relations in Austronesian languages*, Peter K. Austin & Simon Musgrave, eds., Stanford: Center for the Study of Language (CSLI), 90–127.

Austin, Peter K. (2001) Verbs, valence, and voice in Balinese, Sasak, and Sumbawan. In *La Trobe working papers in linguistics*, vol. 11, Bundoora, Victoria: Linguistics Program, La Trobe University, URL `http://hdl.handle.net/1959.9/146325`.

Austin, Peter K. & Simon Musgrave, eds. (2008) *Voice and grammatical relations in Austronesian languages*. Stanford, CA: CSLI Publications.

Blake, Barry J. (1988) Tagalog and the Manila-Mt. Isa axis. In *La Trobe working papers in linguistics*, vol. 1, Bundoora, Victoria: Linguistics Program, La Trobe University,, 77–90.

Blake, Barry J. (1993) Ergativity in the Pacific. In *La Trobe working papers in linguistics*, vol. 6, Bundoora, Victoria: Linguistics Program, La Trobe University,, 19–32.

Blust, Robert (2004) Austronesian nasal substitution: A survey. *Oceanic Linguistics* : 73–148.

Brainard, Sherri (1994) Voice and ergativity in Karao. In *Voice and inversion*, *Typological Studies in Language*, vol. 28, Talmy Givón, ed., Amsterdam/Philadelphia: John Benjamins Publishing Company, 365–402.

Bybee, Joan (2010) *Language, usage and cognition*. Cambridge University Press.

Cartier, Alice (1976) Une Langue a Double Construction Objective et Ergative: l'Indonesien. *La Linguistique* **12**: 99–130.

Cartier, Alice (1979) De-voiced transitive verb sentences in formal Indonesian. In *Ergativity: Towards a theory of grammatical relations*, Frans Plank, ed., New York: Academic Press, 161–183.

Ceña, Resty M. (1979) Tagalog counterexamples to the accessibility hierarchy. *Studies in Philippine Linguistics* **3**(1): 119–124.

Chafe, Wallace L. (1994) *Discourse, consciousness, and time: The flow and displacement of conscious experience in speaking and writing*. University of Chicago Press.

Chung, Sandra (1976a) An object-creating rule in Bahasa Indonesia. *Linguistic Inquiry* : 41–87.

Chung, Sandra (1976b) On the Subject of Two Passives in Indonesian. In *Subject and Topic*, Charles Li, ed., Academic Press, 59–98.

Clynes, Adrian (1995) *Topics in the phonology and morphosyntax of Balinese based on the dialect of Singaraja, North Bali*. Ph.D. thesis, The Australian National University, Canberra, Australia.

Comrie, Bernard (1989) *Language universals and linguistic typology: Syntax and morphology*. Chicago: University of Chicago press, 2 ed.

Cooreman, Ann, Barbara Fox, & Talmy Givón (1984) The discourse definition of ergativity. *Studies in Language* **8**(1): 1–34.

Cooreman, Ann M, Barbara Fox, & Talmy Givón (1988) The discourse definition of ergativity: A study of Chamorro and Tagalog texts. *Studies in Austronesian Linguistics. Athens, Ohio: Ohio University Center for International Studies* : 387–425.

Croft, William (2001) *Radical construction grammar: Syntactic theory in typological perspective*. Oxford University Press.

Cumming, Susanna (1991) *Functional change: The case of Malay constituent order*. 2, Walter de Gruyter.

Cumming, Susanna & Fay Wouk (1987) Is there 'discourse ergativity' in Austronesian languages? *Lingua* **71**(1): 271–296.

Dardjowidjojo, Soenjono (1978) *Sentence patterns of Indonesian*. University of Hawaii Press.

De Guzman, Videa P (1992) Morphological evidence for primacy of patient as subject in Tagalog. *Papers in Austronesian linguistics* **2**: 87–96.

Dixon, R.M.W. (1994) *Ergativity*. Cambridge: Cambridge University Press.

Du Bois, John W. & Sandra A. Thompson (1991) Dimensions of a theory of information flow, unpublished Ms. University of California, Santa Barbara.

Englebretson, Robert (2007) Grammatical resources for social purposes: Some aspects of stancetaking in colloquial Indonesian conversation. In *Stancetaking in discourse: Subjectivity, evaluation, interaction*, Robert Engelbretson, ed., Amsterdam/Philadelphia: John Benjamins Publishing, 69–110.

Evans, Nicholas & Toshiki Osada (2005) Mundari: The myth of a language without word classes. *Linguistic Typology* **9**(3): 351–390.

Ewing, Michael C (2005) *Grammar and inference in conversation: Identifying clause structure in spoken Javanese*. Amsterdam: John Benjamins.

Foley, William A. (1998) Symmetrical voice systems and precategoriality in Philippine languages. In *Workshop on voice and grammatical relations in Austronesian languages*, Brisbane: LFG98 Conference, URL `http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.7.7620`.

Foley, William A. (2008) The place of Philippine languages in a typology of voice systems. In *Voice and grammatical relations in Austronesian languages*, Peter K. Austin & Simon Musgrave, eds., Stanford, CA: CSLI Publications, 22–44.

Gil, David (1994) The structure of Riau Indonesian. *Nordic Journal of Linguistics* **17**(2): 179–200.

Gil, David (2002) The Prefixes di- and N- in Malay/Indonesian Dialects. In *The history and typology of western Austronesian voice systems*, Fay Wouk & Malcolm Ross, eds., Canberra, Australia: Pacific Linguistics, 241–283.

Givón, Talmy (1983) *Topic continuity in discourse: A quantitative cross-language study*, vol. 3. John Benjamins Publishing Company.

Goldberg, Adele (2006) *Constructions at work: The nature of generalization in language.* Oxford: Oxford University Press.

Gordon, Raymond G. (2005) *Ethnologue: Languages of the world.* Dallas, Texas: SIL International, 15 ed.

Helfrich, Oscar Louis (1893) Bijdrage tot de Letterkunde van den Běsěmaher en Sěrawajer in de afdeeling Manna en P.O. Manna (Residentie Bengkoelen). *Tijdschrift voor Indische Taal-, Land- en Volkenkunde, Uitgegeven door het Bataviaasch Genootschap van Kunsten en Wetenschappen* **37**: 65–97.

Helfrich, Oscar Louis (1895a) ERRATA voorkomende in de "Bijdrage" Sěrawajsche en Běsěmahsche spreekworden, spreekwijzen en raadsels. *Bijdragen tot de Taal-, Land-en Volkenkunde* **45**(1): 384–386.

Helfrich, Oscar Louis (1895b) Sěrawajsche en Běsěmansche Spreekwoorden, Spreekwijzen en Raadsels. *Bijdragen tot de Taal-, Land- en Volkenkunde* **45**(1): 1–78.

Helfrich, Oscar Louis (1904) *Bijdragen tot de kennis van het Midden Maleisch: (Běsěmahsch en Sěrawajsch Dialect), Verhandelingen van het Bataviaasch Genootschap van Kunsten en Wetenschappen*, vol. 53. Batavia: Landsdrukkerij.

Helfrich, Oscar Louis (1915) *Nadere Aanvullingen en Verbeteringen op de Bijdragen tot de kennis van het Midden Maleisch (Běsěmansch en Sěrawajsch Dialect) - Lampongsche dwerghertverhalen, Verhandelingen van het Bataviaasch Genootschap van Kunsten en Wetenschappen*, vol. 66. Batavia: Albrecht & Co.

Helfrich, Oscar Louis (1921) *Supplement op de in Deel LXI, 3e en 4e stuk der Verhandelingen Gepubliceerde Nadere Aanvullingen en Verbeteringen op de Bijdragen tot de kennis van het Midden Maleisch (Běsěmansch en Sěrawajsch) (Verschenen in Deel LIII der Verhandelingen), Verhandelingen van het Bataviaasch Genootschap van Kunsten en Wetenschappen*, vol. 63. Batavia: Albrecht & Co.

Helfrich, Oscar Louis (1927a) *Nadere Bijdragen tot de kennis van het Midden Maleisch (Běsěmasch en Sěrawajsch Dialect), Verhandelingen van het Bataviaasch Genootschap van Kunsten en Wetenschappen*, vol. 68. 's-Gravenhage: Nederlandsche Boeken Steendrukkerij V/H H. L. Smits.

Helfrich, Oscar Louis (1927b) Uit de Folklore van Zuid-Sumatra. *Bijdragen tot de Taal-, Land- en Volkenkunde* **83**: 193–315.

Helfrich, Oscar Louis (1933) *Bijdragen tot de kennis van het Midden Maleisch (Běsěmahsch en Sěrawajsch Dialect) Supplement op de "Nadere bijdragen" (1927).* 's-Gravenhage: Martinus Nijhoff.

Himmelmann, Nicholas P. (2002a) Voice in two northern Sulawesi languages. In *The history and typology of western Austronesian voice systems*, vol. Wouk, Fay and Ross, Malcolm, Canberra, Australia: Pacific Linguistics, 123–142.

Himmelmann, Nikolaus P (1991) The Philippine challenge to universal grammar. In *Arbeitspapier, Institut für Sprachwissenschaft, Universität Köln*, vol. 15, Köln: Allgemeine Sprachwissenschaft, Institut für Linguistik, Universität zu Köln, 1–59.

Himmelmann, Nikolaus P (2002b) Voice in western Austronesian: An update. In *The history and typology of western Austronesian voice systems*, Fay Wouk & Malcolm Ross, eds., Canberra, Australia: Pacific Linguistics, 7–16.

Himmelmann, Nikolaus P. (2005) The Austronesian languages of Asia and Madagascar: Typological characteristics. In *The Austronesian languages of Asia and Madagascar, Routledge Family Language Series*, vol. 7, K Alexander Adelaar & Nikolaus P. Himmelmann, eds., New York, NY: Routledge, 110–181.

Himmelmann, Nikolaus P (2008) Lexical categories and voice in Tagalog. In *Voice and grammatical relations in Austronesian languages*, Peter K. Austin & Simon Musgrave, eds., Stanford, CA: Center for the Study of Language (CSLI), 247–293.

Hopper, Paul (1983) Ergative, passive, and active in Malay narrative. *Discourse perspectives on syntax* : 67–88.

Hopper, Paul (1988) How ergative is Malay. *Studies in Austronesian linguistics* (76): 441.

Hopper, Paul J. (1979a) Aspect and foregrounding in discourse. In *Discourse and syntax*, Talmy Givón, ed., New York: Academic Press, 213–241.

Hopper, Paul J. (1979b) Some discourse sources of ergativity. *Working papers in linguistics* **11**: 137–153.

Hopper, Paul J (1985) Ergativity in Western Austronesian: the Malay evidence. *SUNY at Binghamton Manuscript* .

Hopper, Paul J. & Sandra A. Thompson (1980) Transitivity in Grammar and Discourse. *Language* **56**(2): 251–299.

Huang, Shuanfan (2002) The pragmatics of focus in Tsou and Seediq. *Language and Linguistics* **3**(4): 665–694.

Huang, Shuanfan & Michael Tanangkingsing (2011) A Discourse Explanation of the Transitivity Phenomena in Kavalan, Squliq, and Tsou. *Oceanic Linguistics* **50**(1): 93–119.

Kana, Marit Ann (1986) *Grammatical Relations in Bahasa Indonesia*. Ph.D. thesis, Southeast Asia Program, Cornell University.

Kittilä, Seppo (2011) Transitivity typology. In *The Oxford handbook of linguistic typology*, Jae Jung Song, ed., Oxford: Oxford University Press, 346–367.

Kroeger, Paul (1993) *Phrase structure and grammatical relations in Tagalog*. Center for the Study of Language (CSLI).

Lambrecht, Knud (1994) *Information structure and sentence form: Topic, focus, and the mental representations of discourse referents.* Cambridge: Cambridge University Press.

Liao, Hsiu-chuan (2004) *Transitivity and ergativity in Formosan and Philippine languages.* Ph.D. thesis, University of Hawaii at Manoa, Honolulu, HI.

Macdonald, Ross R. & Soenjonjo Darjowidjojo (1967) *A student's reference grammar of modern formal Indonesian.* Washington, D.C.: Georgetown University Press.

McCune, Keith (1979) Passive function and the Indonesian passive. *Oceanic Linguistics* **18**(2): 119–169.

McDonnell, Bradley (2009) A Conservative Vowel Phoneme Inventory of Sumatra: The Case of Besemah. *Oceanic Linguistics* **47**(2): 409–432.

McDonnell, Bradley (2010) *Two passives and a paradigm: Exploring undergoer voice in Besemah.* Master's thesis, Arizona State University, Tempe, AZ.

McDonnell, Bradley (2013a) Causative/applicative syncretism: Besemah Malay -kah. In *The 87th Annual Meeting of the LInguistics Society of America*, Boston, MA.

McDonnell, Bradley (2013b) Noun-modifying clause constructions in Besemah. In *Association for Linguistic Typology 10th Biennial Conference*, Leipzig, Germany.

McDonnell, Bradley (2013c) Roadblocks in the grammaticalization highway: When phonology gets in the way. In *Proceedings of the Annual Meeting of the Berkeley Linguistics Society*, vol. 39.

McDonnell, Bradley (2014) Acoustic correlates of stress in Besemah, unpublished ms. University of California, Santa Barbara.

Mithun, Marianne (2008) Does passivization require a subject category? In *Case and grammatical relations: Studies in honor of Bernard Comrie*, Greville G. Corbett & Michael Noonan, eds., Amsterdam: John Benjamins, 211–240.

Musgrave, Simon (2000) Pronouns and morphology: Undergoer subject clauses in Indonesian. In *Yearbook of Morphology 2000*, Springer, 155–186.

Musgrave, Simon (2001) *Non-subject arguments in Indonesian.* Ph.D. thesis, University of Melbourne, Department of Linguistics and Applied Linguistics.

Naylor, Paz B (1995) Subject, topic, and Tagalog syntax. *Subject, voice and ergativity* : 161–201.

Pastika, I Wayan (1999) *Voice selection in Balinese narrative discourse.* Ph.D. thesis, The Australian National University, Canberra, Australia.

Payne, Thomas E (1982) Role and reference related subject properties and ergativity in Yup'ik Eskimo and Tagalog. *Studies in Language Groningen* **6**(1): 75–106.

Payne, Thomas E (1994) The pragmatics of voice in a Philippine language: Actor-focus and goal-focus in Cebuano narrative. In *Voice and inversion, Typological Studies in Language*, vol. 28, Talmy Givón, ed., Amsterdam/Philadelphia: John Benjamins Publishing Company, 317–64.

Rafferty, Ellen (1982) *Discourse structures of the Chinese Indonesian of Malang.* Jakarta, Indonesia: Badan Penyelenggara Seri Nusa, Universitas Atma Jaya.

Reid, Lawrence A. & Hsiu-chuan Liao (2004) A brief syntactic typology of Philippine languages. *Language and Linguistics* **5**(2): 433–490.

Ross, Malcolm (2002) The history of western Austronesian voice and voice-marking. In *The history and typology of western Austronesian voice systems*, Fay Wouk & Malcolm Ross, eds., Canberra, Australia: Pacific Linguistics, 17–62.

Schachter, Paul (1976) The subject in Philippine languages: Topic, actor, actor-topic, or none of the above actor, actor-topic, or none of the above. *Subject and topic* : 491–518.

Schachter, Paul (1977) Reference-related and role-related properties of subjects. *Syntax and semantics* **8**: 279–306.

Schachter, Paul (1996) *The subject in Tagalog: still none of the above*, vol. 15. UCLA.

Sneddon, James N. (1996) *Indonesian: a comprehensive grammar.* London: Routledge.

Tchekhoff, Claude (1978) *Aux fondements de la syntaxe: l'ergatif.* Paris: Presses Universitaires du France.

Tchekhoff, Claude (1980) The Economy of a Voice-Neutral Verb. An Example in Indonesian. In *Austronesian Studies: Papers from the Second Eastern Conference on Austronesian Languages*, 71–81.

Thompson, Sandra A & Paul J Hopper (2001) Transitivity, clause structure, and argument structure: Evidence from conversation. In *Frequency and the Emergence of Linguistic Structure, Typological Studies in Language*, vol. 45, Joan L. Bybee & Paul J. Hopper, eds., Amsterdam: John Benjamins Publishing Company, 27–60.

Vendler, Zeno (1957) Verbs and times. *The Philosophical Review* **66**(2): 143–160.

Verhaar, John WM (1983) Ergativity, accusativity, and hierarchy. *Sophia Linguistica. Working Papers in Linguistics Tokyo* (11): 1–23.

Verhaar, John WM (1984) Towards a description of contemporary Indonesian: preliminary studies Part II. *NUSA, Linguistic Studies of Indonesian and other Languages in Indonesia* **19**.

Verhaar, John WM (1988) Syntactic ergativity in Indonesian. In *Studies in Austronesian linguistics, Monographs in International Studies, Southeast Asia Series*, vol. 1988, Richard McGinn, ed., Athens, Ohio: Center for International Studies, Ohio University, 347–384.

Wechsler, Stephen & I Wayan Arka (1998) Syntactic ergativity in Balinese: an argument structure based theory. *Natural Language & Linguistic Theory* **16**(2): 387–442.

Wolff, John U (1996) The development of the passive verb with pronominal prefix in Western Austronesian languages. *Reconstruction, classification, description: Festschrift in honor of Isidore Dyen* : 15–40.

Wouk, Fay (1989) *The Impact of Discourse on Grammar: Verb Morphology in Spoken Jakarta Indonesian.* Ph.D. thesis, UCLA.

Wouk, Fay (1996) Voice in Indonesian Discourse and Its Implications for Theories of the Development of Ergativity. *Studies in Language* **20**(2): 361–410.

Wouk, Fay (1999) Sasak is Different: A Discourse Perspective on Voice. *Oceanic Linguistics* **38**(1): 91–114.

Wouk, Fay (2004) Nasal transitive prefixes in spoken Jakartan Indonesian. *NUSA* **53**: 21–28.

Wouk, Fay & Malcolm Ross, eds. (2002) *The history and typology of western Austronesian voice systems.* 518, Pacific Linguistics.

A Independent variables

In the the appendix, I outline the properties that I plan to code in the Besemah corpus. I will consider all of these properties, but will first focus on those which are marked with '*'.

Table 2: Independent variables of the formal properties of the verb and its arguments

Variable	Levels
AGENT PRESENCE*	yes, no
AGENT BOUNDEDNESS*	bound, free, NA
AGENT DEFINITENESS*	yes, no (marked by determiner/demonstrative)
AGENT POSITION*	preverbal, post-verbal
AGENT ANIMACY*	animate,inanimate
AGENT LENGTH*	0, 1, 2, 3 (# of characters)
AGENT TOPICALIZED*	yes, no (marked by topicalizer ame/amu)
PATIENT PRESENCE*	yes, no
PATIENT BOUNDEDNESS*	bound, free, NA
PATIENT DEFINITENESS*	yes, no (marked by determiner/demonstrative)
PATIENT POSITION*	preverbal, post-verbal
PATIENT ANIMACY*	animate,inanimate
PATIENT LENGTH*	0, 1, 2, 3 (# of characters)
PATIENT TOPICALIZED*	yes, no (marked by topicalizer ame/amu)
WORD ORDER*	AVP, VPA, NA
TENSE/ASPECT*	unmarked, progressive, perfective, imperfective
SUBORDINATION*	subordinate-clause, main-clause

Table 3: Independent variable of the verbal semantics (Vendler 1957)

Variable	Levels
ROOT CATEGORY*	noun, verb
VERB ROOT TRANSITIVITY*	intransitive, transitive, ditransitive
VERB ROOT SEMANTICS*	states, activities, accomplishments, achievements
VERB TRANSITIVITY*	intransitive, transitive, ditransitive

Table 4: Independent variables of information flow (Du Bois & Thompson 1991)

Variable	Levels
IDENTIFIABILITY*	*identifiable, non-indentifiable*
IDENTIFIABILITY PATHWAY	*previous mention, proposition, participant, presence, anchoring, association, world knowledge*
ACTIVATION STATE*	*given, new*
GENERALITY*	*particular, general*
DISCOURSE REFERENTIALITY*	*tracking, non-tracking*

Table 5: Independent variables of discourse transitivity (Hopper & Thompson 1980, Thompson & Hopper 2001)

Variable	Levels
KINESIS	*1 (high kinesis) to 3 (low kinesis)*
ASPECT	*telic, atelic*
PUNCTUALITY	*punctual, non-punctual*
AGENT VOLITIONALITY	*volitional, non-volitional*
AFFIRMATION*	*affirmative, negative*
MODE*	*realis, irrealis*
AGENCY	*1 (low in potency) to 3 (high in potency)*
PATIENT AFFECTEDNESS*	*1 (patient not affected) to 3 (patient highly affected)*
PATIENT INDIVIDUATION*	*1 (patient not individuated) to 3 (patient highly individuated)*

Table 6: Independent variables of topic continuity (Givon 1983)

Variable	Levels
AGENT REFERENTIAL DISTANCE*	*1 to 20 (# of IUs intervening between last mention/evocation)*
PATIENT REFERENTIAL DISTANCE*	*1 to 20 (# of IUs intervening between last mention/evocation)*
AGENT PERSISTENCE	*# of IUs an agent persists*
PATIENT PERSISTENCE	*# of IUs a patient persists*
AGENT INTERFERENCE	*interference, non-interference*
PATIENT INTERFERENCE	*interference, non-interference*
AGENT PREVIOUS MENTIONS*	*# of previous mentions*
PATIENT PREVIOUS MENTIONS*	*# of previous mentions*
AGENT EVOKED	*scored 0–1*
PATIENT EVOKED	*scored 0–1*