How report verbs become quote markers and complementisers[☆]

Marian Klamer*

Taalkunde/ATW, Vrije Universiteit Amsterdam, P.O. Box 7161, 1007 MC Amsterdam, The Netherlands

Abstract

In many languages across the world, verbs reporting speech, thoughts and perceptions (also referred to as quotative verbs) grammaticalise into quote markers and/or complementisers. This paper analyses the change of the items *kua* and *fen* in the Austronesian languages Tukang Besi and Buru, as originally full lexical 'report' verbs that became open to reinterpretation as grammatical items after having undergone 'semantic bleaching'. It is proposed that the 'semantic bleaching', which crucially involves loss of argument structure, is caused by a mismatch between linguistic levels – here between surface syntax and lexical argument structure. The mismatch involves a violation of universal constraints on 'Semantic Transparency' and 'Structural Simplicity', and results in a reduced lexical representation of the report verb as a predicate without arguments. The multifunctional, polysemous character of this 'grammaticalised' item is now a consequence of its interaction with particular surface syntactic constructions. In other words, 'V to C' grammaticalisation is a structurally determined variable interpretation of a lexically impoverished item, and does not involve a change in category (labels) (contra Harris and Campbell, 1995:63; Heine and Reh, 1984: 37–38; see also Haspelmath, 1998: 327–328). This view of grammaticalised verbs as lexical forms with reduced argument structure may be extended to other areas of verb-grammaticalisation. The

similar path of grammaticalisation of report verbs across languages is explained by proposing
a list of structural characteristics (of syntax and discourse) that appear to be relevant in allow-
ing the grammaticalisation to take place. Genetically related languages may diverge because
they differ in one (or more) of those characteristics: the report verb in Kambera, a language
closely related to Buru and Tukang Besi, did not grammaticalise because of a different sur-
face constituent order. © 2000 Elsevier Science B.V. All rights reserved.

Keywords: Verb grammaticalisation; Quotative/report construction

1. Introduction

In this paper I discuss the grammaticalisation of constructions that report speech,
thoughts and perceptions, with a focus on the Austronesian languages Kambera,
Buru and Tukang Besi. Kambera and Buru are both classified as Central Malayo-
Polynesian languages. Kambera has 150,000 speakers, is spoken on the island of
Sumba in Eastern Indonesia, and described in Klamer (1998). Buru has 43,000
speakers, is spoken in Central Maluku, Indonesia, and described in Grimes (1991).
Tukang Besi is classified as Western Malayo-Polynesian, has 200,000 speakers, is
spoken in Southeast Sulawesi, and described in Donohue (1995). The grammatical
elements *kua* (Tukang Besi) and *fen* (Buru) function as elements linking two clauses
together:[1]

(1) *Tukang Besi*
     No-'ita-'e   kua   no-kanalako te   osimpu
     3R-see-3Obj   KUA   3R-steal   Core   young coconut
     'She saw that he had stolen the coconut'
(2) *Buru*
     Sira   em-tako   fen   sira dapak   eflali
     3p   Stat-fear FEN   3p   get   beat
     'They were afraid that they would be beaten'

I refer to this 'clause-linking' function as the 'complementiser' function of
kua/fen. There is no syntactic difference between the clauses preceding and follow-
ing *fen/kua* – both clauses are finite and both contain an overt subject marker. In

---

[1] The Buru data used in this paper are from Grimes (1991: 224, 396–399, 407–409, 425–429, 472,
531). The Tukang Besi data are from Donohue (1995), esp. Section 12.8 and Chapter 16. In the
Buru/Tukang Besi examples I have used the original glosses except for the elements *kua* and *fen*, these
are not glossed. Glossing conventions in the Buru/Tukang Besi data: Core = a case-marking article for
a "core argument that is within the clause but not in nominative case" (Donohue, 1995: 46), Dat =
Dative, Dist = Distant, Gen = Genitive, Irr = Irrealis, Loc = Locative, NL = Nominaliser, Nom = Nom-
inative, Obl = Oblique, P/pl = Plural, Prf = Perfective, R = Realis, S/s = Singular, Seq = Sequential,
Stat=Stative marker. Glosses in the Kambera data: A = Accusative, Art = Article, Cnj = Conjunction,
Cont = Continuative aspect, D = Dative, Dei = Deictic element, Der = Derogatory marker, Exc = Excla-
mation, Exist = Existential marker, G = Genitive, Impf = Imperfective, Loc = Location, Mod = Modal
marker, N = Nominative, Neg = Negation, p = plural, Prf = Perfective, s = singular.

complex sentences, Buru and Tukang Besi make extensive use of simple clause juxtaposition, and clause connecting elements such as *kua/fen* are often optional. When *kua* and *fen* follow main verbs that report speech, thought and physical perception ('report verbs'), they have a complementiser function. In addition, both *kua* and *fen* function as quote markers, while Tukang Besi *kua* can also be interpreted as a marker of logophoric contexts, and Buru *fen* has an additional verbal function.

In this paper I first trace the common source of the various synchronic incarnations of *kua/fen*. Because neither Buru nor Tukang Besi has historical documentation, I use synchronic comparative evidence from a genetically related third language, Kambera. I argue that both *kua* and *fen* were originally verbs used in 'quotative constructions' reporting speech, thoughts and perceptions. Kambera still displays the pattern that the other two languages have grammaticalised.

In many respects, the development of *kua/fen* displays the typical pattern of verbs grammaticalising into quote markers and/or complementisers, a change that has been observed to occur in numerous languages across the world (see, for example, the overview in Lord, 1993: Ch. 7). The goal of this paper is therefore to go beyond a descriptive reconstruction of the Austronesian facts, and present a general account of this type of grammaticalisation.

It is an important characteristic of grammaticalised items that, synchronically, they are often multifunctional items that belong to different lexical or functional word classes. Also, grammaticalised items differ from full lexical items because they have undergone 'semantic bleaching' (Hopper and Traugott, 1993). I analyse the 'semantic bleaching' of report verbs as involving loss of argument structure, and argument loss is the result of a 'mismatch' between linguistic levels – in this case between surface syntax and lexical argument structure.[2] The mismatch involves a violation of two constraints on form-function mapping: 'Semantic Transparency' and 'Structural Simplicity'. It results in a lexical representation of the report verb as a predicate without arguments. The multifunctional, polysemous nature of this 'grammaticalised' item is then a direct consequence of its interaction with surface syntax. For example, when it precedes an intonational break and an actual quote, it is interpreted as a quote marker, but when it precedes the second clause of a sentence without being separated from it by an intonational break, it is interpreted as a complementiser. In other words, 'V to C' grammaticalisation is analysed here as not involving a change in category (labels) (contra Heine and Reh, 1984; Harris and Campbell, 1995).

Not surprisingly, the genetically related languages Buru and Tukang Besi share certain structural typological characteristics. These shared structural characteristics explain why the grammaticalisation process of report verbs into complementisers is so similar in both languages – not their genetic relatedness as such. Genetic related

---

[2]   The mechanism of argument structure loss is a widespread source of grammaticalisation. It has often been observed (e.g. Lightfoot, 1979; Roberts, 1985) that modal verbs or raising verbs with modal force, which have an impoverished argument structure, originate from verbs with a full argument structure. Examples include modal *get* in English (*John gets to go to the movies*), and raising verbs such as *promise* (*It promises to be a beautiful day*). See also notes 16 and 17.

ness does not guarantee similar paths of language change: we will see that the report verb in the third related language Kambera did not undergo the grammaticalisation process. This is explained by the fact that the Kambera report constructions have a different (surface) constituent order.

The paper is structured as follows. The rich and productive use of report constructions in Kambera is described in Section 1 and contrasted with the more fossilised use of similar constructions with Buru *fen* in Section 2 and Tukang Besi *kua* in section 3. In Section 4 I propose a general account of this type of grammaticalisation. In Section 5 I point out how the view that grammaticalised verbs are lexical forms with reduced argument structure may be extended to other areas of verb-grammaticalisation, and I then summarise the findings of the paper.

2. Kambera report constructions

Kambera is a head-marking language; verbal arguments are commonly marked on the verb by pronominal clitics, and coreferent NPs are normally optional. The regular form for Patients is an accusative clitic, as in (3). Beneficiaries/Recipients are regularly marked with a dative clitic, as in (4) (cf. Klamer, 1998, for details):[3]

(3) Da- ngàndi -**ya**   na   uhu
    3pN- take   -3sA   ART rice
    'They take the rice'

(4) Da- ngàndi -**nya** na   uhu i   Ama
    3pN- take   -3sD   ART rice ART father
    'They bring father the rice'

The Kambera quotative verb *wà* is an intransitive root verb. The speaker is marked by a genitive enclitic:

(5) "...", wà -nggu
    "...", say -1sG
    '"...", I said'

Wà can be morphologically derived with an applicative suffix *-ng*, which adds an 'addressee' argument. This argument is expressed by a dative object clitic. An illustration of *wà-ng* used as 'talk to' is:[4]

(6) E,   wà -nggu -nya$_j$ [na   ama   -mu]$_j$!
    EXC say -1sG -3sD   ART father -2sG
    'Hey, I was talking to your father!'

---

[3] Patients are marked with the (prenasalised) dative forms when the citation form of the verb ends in a nasal, as in (7), where the citation form of the verb is *píng* 'know'.

[4] The applicative nasal suffix disappears for morpho-phonological reasons. For detail I refer to Klamer (1998).

There are several arguments to analyse *wà* as intransitive. Firstly, Kambera clausal clausal complements (such as nominalised clauses) are crossreferenced with object clitics on the main verb, as in (7).

(7) Nda   ku- pí   **-nya** na   laku-mu
    NEG 1sN- know -3sD   ART   go-2sG
    'I didn't know that you'd gone'

Quotes, however, are simply juxtaposed to the clause containing the quotative verb (simple *wà*, or derived *wà-ng*); they are *never* crossreferenced with a pronominal element on the quotative verb. Secondly, *wà* cannot appear in passive-like structures, unlike normal transitive verbs (cf. Klamer, 1996, 1998, 1999). Also, there are constituent order differences between quotes and other verbal complement clauses: unmarked constituent order is VO, but quotes exceptionlessly *precede* the quote verb, and would thus instantiate OV order if the quote were a verbal complement. Furthermore, coordinating conjunctions such as *hi* 'then, thus, and', *ka* 'so that', *ba* 'as, while, because' and *jàka* 'if' appear optionally between the quote and the quotative verb, suggesting that we are dealing with two coordinated independent clauses. In (8) the conjunction *hi* is illustrated; this conjunction is optional:

(8) 'Ai   Umbu,   ka   nda   u-mila-ngga   nú   eti'   **hi**   wà-na-nya
    EXC Lord   CNJ   NEG   2sN-be.poor-1sD   DEI   liver CNJ   report-3sG-3sD
    '"Oh sir, if you would take pity on me," he said to him'
    (lit. 'Wouldn't you have a poor liver for me')

Finally, the analysis of the quote verb *wà* as intransitive converges with the crosslinguistic observations about 'say' verbs of Munro (1982) and De Roeck (1994). In De Roeck's sample (representative for the world's language families), 37,5% of the verbs of behave intransitively, and 10% behave transitively only with an Addressee argument (like Kambera *wà-ng*). Only 47,5% allow for the quote to be treated as the verbal complement. In other words, typologically, the Kambera-type 'non-complement' quotes are just as common as the English type 'complement' quotes.

The root verb *wà* (and, of course, its applicative derivation *wà-ng*) is used to report direct as well as indirect speech. In Kambera there is no syntactic distinction between the two, as illustrated in (9):

(9) "Mài-kai-wa"   wà-na-nggai
    come-2pA-HORT   report-3sG-2pG
    'He says that you (pl) should come/ He tells you (pl) to come'
    '"Come," he says to/tells you'

However, different pronominal reference strategies may be used to indicate distinct speaker perspectives. This is illustrated in (10), where the quote is *na-lua haromu* 'she goes tomorrow'. The subject pronominal (3 sg) in the quote does not

refer directly to the speaker (1 sg), signalling that this is an indirect speech report. But the syntax of the sentence (verbal forms, aspect marking etc.) does not differ from direct quotes.

(10) Na-lua haromu   wà-na-ngga
     3sN-go tomorrow report-3sG-1sD
     'She told me that she is leaving tomorrow'
     'She told me "she leaves tomorrow"'

Apart from speech act contexts, the verb *wà* is also employed to report mental perception. In such contexts the 'quote' is a thought rather than an actual speech act, and *wà-ng* may be translated as e.g. 'know', 'think', or 'realise':

(11) Nda   na-tanda-a-ya   una   na   bai...
     NEG 3sN-know-MOD-3sA DEI.3s ART DER
     'She didn't recognize him, that (man) ...
     jia   na   lei-nggu   amang nda wà-na   mbu -pa   una
     EXIST ART husband-1sG earlier NEG say-3sG also-IMPF DEI.3s
     she didn't even realise he was her former husband'
     (lit. 'she didn't even say, "he used to be my husband"')

Finally, the verb is used in constructions with ideophones. Ideophonic roots express sounds, motions or sights, i.e. various kinds of physical perception:

(12) *Kambera ideophonic roots*

Sounds		Motions		Sights	
ngùru	'murmur'	*yidi*	'shiver' (in dislike)	*jila*	'glimmer/flash'
hèri	'tearing noise'	*wàdi*	'blink'	*bila*	'light/brightness'
tòru	'rattle'	*ngàdu*	'nod' V: agree		
ndùru	'roll' (thunder)	*linji*	'jump'		
pàka	'smack'	*nggidi*	'shiver' (of cold)		
mbùku	'snap/tap'				
tiku	'creak/click'				

Ideophonic roots appear as quotes in a construction with the verb *wà* as illustrated in (13). The subject of the clause is marked with the genitive clitic attached to *wà*, while the ideophonic root itself expresses the perceived state of affairs and is found in the position that would normally be occupied by the quote. In this way, the perception of states of affairs is given a vivid, lively and direct sense.[5]

(13) Mbùtu wà-na   tuna nú, na-puru   nuna nú
     thud   say-3sG thus DEI 3sN-descend DEI.3s DEI
     'Thud! it did and he climbed down there'

---

[5]   Other constructions with ideophonic roots are: reduplications and verbal derivations with the circumfix *ka—k*, cf. Klamer (1998) for more details.

We conclude that in Kambera the report construction is used to express speech acts and (in)direct speech reports, mentally perceived events ('realise') and physically perceived events (the constructions with ideophones).

Across the world's languages, 'quotative' or 'report' constructions are described as expressing speech acts and mental/physical perception (cf. e.g. Reesink, 1993; De Vries, 1990, for Papuan; Adelaar, 1990, for Andean languages; see Lord, 1993, for references on African, Southeast Asian and pidgin/creole languages). The similar marking of speech and cognitive acts has been explained by considering both of them a kind of speech – cognitive activities are a type of 'inner speech' (Vygotsky, 1962). However, Kambera is one of the languages that shows that not only mental events such as thinking are expressed in quotative constructions, but also physical perceptions of motions, sounds and visions.[6] Obviously, perceived events are not necessarily a type of inner or outer 'speech'. What the three uses of the quotative construction have in common, however, is that all of them *report* something: words, thoughts or perceptions, and speech is just one of the possible types of events that may be reported on. This means that a verb such as the Kambera verb *wà* must have a more general semantics: [REPORT], and 'quotative constructions' are a particular type of 'report' constructions.

Crosslinguistically it is not unusual to find report constructions used for mental and physical perception as well as quotation (cf. the references in Harris and Campbell, 1995; Lord, 1993). Within the Malayo-Polynesian branch of the Austronesian language family, however, Kambera seems rather exceptional in its wide use of this construction. Sources on individual Malayo-Polynesian languages such as Muna (Van den Berg, 1985), Leti (Van Engelenhoven, 1995), Karo Batak (Woollams, 1996), Tetun (Van Klinken, 1997) and Savu (Walker, 1982) suggest this. However, in this paper I discuss evidence that at least two other Malayo-Polynesian languages, Buru and Tukang Besi, have employed a report construction with functions similar to the Kambera one. I also explain why in Buru and Tukang Besi the construction became reinterpreted, while in Kambera it did not.[7]

2. Buru report constructions

The Buru element *fen(e)* occurs with a similar range of functions as Kambera *wà*. It is supposedly related to the Proto-Malayo Polynesian form *bener* '(be) true' (Grimes, 1991: 134) and is used as a verb meaning 'think, say, affirm' (Grimes,

---

[6] The Niger-Congo language Engenni (Lord, 1993: 192–197; Thomas, 1978) would be another example.
[7] The verb *wà(-ng)* itself has become reanalysed to some extend, resulting is 'frozen' morphological properties such as the following: (i) Kambera verbs commonly allow their subject to be marked by a nominative proclitic or a genitive enclitic, but the subject of *wà(-ng)* may only be genitive (ii) *wà(-ng)* is the only verb where the subject marker is closer to the verb than the mood clitics, i.e. the mood clitics follow rather than precede the subject marker. This may indicate that the genitive subject marking clitic is on the way to being reinterpreted as an inflectional ending on the verb. For discussion see Klamer (1998).

1991: 133, 134). Grimes spells it as either *fen* or *fene*. Buru has other speech-act verbs, such as *prepa* 'speak', *erei* 'refuse', *siu* 'order', *enika* 'ask', *fisara* or *epsara* 'speak formally' (< Sanskrit *bicara*), and *storita* 'tell a story' (< Sanskrit *ceritera*). These verbs all seem semantically more particular than *fene*.

The various functions of *fen* can be described as follows. First there is a two-way distinction based on the presence and absence of an intonational break following *fen*. This is illustrated in the sentences below. The first *fene* in (14) is separated from the next clause by a pause (indicated by the comma). This *fene* can be interpreted as either a verb or a quote marker. The second *fen* is not phonologically separated from the next clause and may be interpreted as a complementiser.

(14) Da prepa la   yako **fene**, "Ku enika ama-n   dii
     3s say   Dat 1s   FEN 2s   ask   father-Gen Dist
     'He spoke to me (saying:) / He said to me (quote:), "Ask father
     fen   ma iko leu-k   fi   doo?"
     FEN 1pl go precede-k LOC where
     where we should go first"' (Grimes, 1991: 407/426)

Grimes (1991: 134) notes that "[speech verbs] tend to take clausal complements". It seems that this is the case for all the Buru speech verbs except *fene*, because the grammar does not contain a sentence where *fene* is the *only* speech verb, followed by a clausal complement (nominalised[8] or not, cf. (31)), and neither does it contain a clause with the single verb *fene* followed by a nominal complement. *Fene* is also not present in the verbal taxonomy of Buru (Grimes, 1991: Ch. 7). It is thus unclear whether verbal *fene* would be transitive or intransitive. What *is* clear is that its morphological and syntactic possiblities of *fene* are very much reduced: it does not take additional morphology (Grimes, 1991: 134) and – hence – does not appear in passive constructrions either. Lacking positive evidence that the verb has a complement, I assume that it is intransitive.

In (15a) *fen* is separated from the quote by an intonational break, in (15b) it is not. In this way Buru makes an intonational distinction between between direct and indirect speech, but observe that there is no morpho-syntactic distinction between the two:

(15) a. Da prepa fen,   "Sira rua   kaduk"
        3s say   FEN 3p   two arrive
        'She said, "The two of them came"' (Grimes, 1991: 396)
     b. Da prepa fen   sira rua   kaduk
        3s say   FEN 3p   two arrive
        'She said that the two of them came' (Grimes, 1991:396)

In situations where there is an intonational break, we may further subdivide the uses of *fen* based on whether it stands alone as an independent 'speech verb' or

---

[8] Buru nominalised clauses can function both as objects and as subjects of transitive verbs (Grimes, 1991: 428–429).

whether it occurs with another verb. The first case is illustrated in (16)–(17). Here, *fene* is used verbally to introduce reported speech and is separated from the reported speech clause by an intonational break. Note that in both sentences the subject is a full NP.

(16) Titas.boti   suba   dii,   fene,   "Kae em-kua   naa?"
     mouse.white cross threshold DIST   say   2s   STAT-why PROX
     'The white mouse showed up and said, "What's the matter with you?'
     (Grimes, 1991: 533)

(17) Nak   ana-t   fene,   "Ng-ina ,   nau   dah.olo"
     3sPoss child-NOM say   1sVOC-mother 1sPoss bunch.head
     'Her child said, "Mother, the hand (i.e. of bananas) at the top of the stalk is for me"'(Grimes, 1991: 531)

The second case, where *fen* appears together with another verb is when it co-occurs with verbs belonging to one of the following semantic classes: verbs of speech, as in (18), or verbs expressing physical perception and mental perception, as in (19). In such contexts, *fen* functions as a complementiser, and its speech verb and quote marking interpretation is absent.

(18) Speech-act verbs occurring with *fen*:
     prepa 'speak, say', *erei* 'refuse', *siu* 'order', *enika* 'ask'

(19) Mental and physical perception verbs occurring with *fen*: *odo* 'think', *tewa* 'know, understand, be able to', *odo* 'think, imagine', *fasa* 'decide', *nanu-k* 'think, understand', *tele-k* 'understand (meaning, significance, implications)' (Grimes, 1991: 133), *kita* 'see'

The following sentences illustrate the complement function of *fen* following verbs of speech. Note that the referential function of the pronominals in the second clause can be different. We return to this below.

(20) **Da** prepa   fene **ringe** mata haik
     3s   speak/say FEN 3s   die   PRF
     'He$_i$ said that he$_j$ was already dead' (Grimes, 1991: 133)

(21) **Ringe** prepa   fen **da** moho
     3s   speak/say FEN 3s fall
     'He$_i$ said that he$_i$ fell' (Grimes, 1991: 224)

(22) Sera erei   fen   du eptea fi   dii
     3p   refuse FEN 3p sit   LOC DIST
     'They refused, saying they would stay here' (Grimes, 1991: 426)

(23) [...], tu   sira enika fen, 'Kae kadu-k naa   la   maksud sapa-n?'
          with 3p ask   FEN 2s   come-k PROX IRR intent   what-GEN
     '[...], but they ask, "Your coming here is for what purpose?"' (Grimes, 1991: 537)

(24) Man.wae   fehu-t   sa   kaduk, du kaba fen   'bebe lau'
     bird.water new-NOM one come-k 3p call FEN duck sea (Malay)
     'A new water-bird came, and they called it/it was called "bebe lau"' (Grimes,
     1991: 364)

Fen is not obligatory with all speech verbs:

(25) ... petu du **fal.ngaa**-i-k   fena   di   Wae Katin
     SEQ   3p throw.name-LOC-k village DIST Water Pandanus
     '... so they named the village/the village was named Pandanus Water'
(26) Gofo-t   **prepa**,   "Gam dii,   do,   kae namu-k luke-n"
     turtle-NOM speak/say SIM   DIST PAUSE 2s   2sPoss-k tip-GEN
     'The turtle said, "Well, if that's how you want it, you take the young leaves at
     the top"' (Grimes, 1991: 544)

The following sentences illustrate the complement function of *fen* following verbs of
mental and physical perception.

(27) Sira em-tako fen   sira dapak eflali
     3p   Stat-fear FEN 3p   get   beat
     'They were afraid that they would be beaten'
(28) Da odo   fene ringe   mata haik
     3s think FEN 3s   die   PRF
     'He$_i$ thought that he$_j$ was already dead' (Grimes, 1991: 133)
(29) Ya tewa fen   ringe iko haik
     1s know FEN 3s   go   PRF
     'I know that he has already left' (Grimes, 1991: 134)
(30) Ya kita fen   da iko haik
     1s see   FEN 3s go Prf
     'I saw that he had already left'

(The restricted distribution of *fen* is also illustrated in (37) and (38), where it cannot
co-occur with an activity verb like *iko* 'go'.)
     Juxtaposition with *fen,* as in (31a), contrasts both with juxtaposition without an
overt clause linker (31b), and with subordination by nominalisation of a clause (31c):

(31) a. Sira kita fen   da iko
         3p   saw FEN 3s go
         'They saw that he left' (Grimes, 1991: 429)
     b. Sira kita da iko
         3p   saw 3s go
         'They saw he left' (Grimes, 1991: 429)
     c. Sira kita nak   en-   yiku -t
         3p   see   3sPoss Abstract marker- go   -Nominaliser
         'They saw his going' (Grimes, 1991: 429)

In its complementiser function, *fen*[9] contrasts with the modal complementiser *la* 'Irrealis', as illustrated in (32) and (33):

(32) a. Sira erei   fen   du eptea fi   dii
        3p   refuse FEN   3p sit   Loc Dist
        'They$_i$ refused, (saying) they$_i$ would stay here'
     b. Sira erei   la   du eptea fi   dii
        3p   refuse IRR   3p sit   Loc Dist
        'They$_i$ refused to stay here'
(33) a. Da prepa fen   ringe iko
        3s say   FEN   3s   go
        'He$_i$ said that he$_j$ went (=left)'
     b. Da prepa   la   ringe iko
        3s say IRR 3s   go
        'He$_i$ said that he$_j$ should go'

Grimes describes the contrast between the two complementisers as follows: *fen* introduces a clause with a "descriptive (indicative)" sense, while *la* indicates a "manipulative (jussive)" sense (Grimes 1991: 426).[10]

In Buru, pronominal clitics (such as *da*) are anaphoric elements, whereas full pronouns (such as *ringe*) are referentially free. Stirling (1993) describes how in switch-reference and logophoric systems distinct subject pronouns in a dependent clause indicate same or different reference with the subject of a controlling clause.

In Buru, switch-reference is marked by employing the distinction between anaphoric and referentially free pronominals. Both clauses have an identical subject referent when the first clause has a full pronoun (e.g. *ringe* 's/he' in (34)) and the second clause a pronominal clitic subject (*da* in (34)), or when both clauses have pronominal clitic subjects, as in (35). (With identical referents, the second pronominal can even be omitted altogether.) Using a full pronoun in the second clause encodes a shift to a different (or reintroduced) referent (Grimes, 1991: 152–153, 255), as in (36).

(34) Ringe odo   fene da mata haik
     he   think FEN 3s die PRF
     'He$_i$ imagined that he$_i$ was already dead'[11]
(35) Da odo   fene da mata haik
     3s think FEN 3s die PRF
     'He$_i$ imagined that he$_i$ was already dead' (Grimes, 1991: 152)

---

[9]   Grimes glosses *fen(e)* as Realis marker, "marking the factivity of the matrix verb, rather than the non-factivity of the complement." (Grimes, 1991: 224, note 15).
[10]   Buru also has a locative preposition *la*, which is also used to mark dative and benefactive arguments (Grimes, 1991: 257). The irrealis complementiser may be derived from this preposition, though synchronically there are functional and distributional differences.
[11]   To show the contrast with the next sentence more clearly, this example was reconstructed on the basis of analogous sentences in Grimes (1991).

(36) Da odo fene ringe mata haik
    3s think FEN s/he die PRF
    'He$_i$ imagined that he$_j$ was already dead' (Grimes, 1991: 152)

What we witness here is an incipient switch-reference or logophoric system. Stirling (1993) shows that such systems may arise as the result of the grammaticalisation of report constructions. Indeed, below we will see that in Tukang Besi, constructions with the (original) speech verb *kua* also have logophoric characteristics. In Tukang Besi, however, the presence/absence of the complementiser is relevant for the reference switch, whereas in Buru this is not the case.

The distribution of the complementiser *fen* is restricted by the semantics of the verb which precedes it, while that of the other complements is more free. Activity verbs such as *iko* 'go' and *oli* 'return' can in general be followed by complementisers such as *la* 'Irrealis' and *petu* 'Sequential'. *Fen*, however, does not occur in such contexts:

(37) Da iko **la** / *fen da kaa
    3s go Irr   3s eat
    'He went to eat'
(38) Da iko, **petu** / *fen da kaa
    3s go Seq   3s eat
    'He went and (then) he ate'

We can explain the synchronic distributional restrictions on *fen* if we assume that historically, Buru (like Kambera) employed quotative constructions to report words, thoughts and perceptions. The proposed development of *fen(e)* is thus that the report verb *fen(e)* has developed an alternative interpretation as a quote marker *fen*, in which case *fen* is separated from the actual quote by an intonational break and has no overt subject.

In contexts where *fen* is preceded by another verb which reports words, thoughts or perceptions, it has developed a complementiser interpretation, and is used to link two clauses together without an intonational break. From Grimes (1991) it appears that *fene* is the only Buru complementiser with a clearly verbal origin.

In the analysis of the grammaticalisation of report verbs in section 4 I propose that the absence of overt subject marking is a surface syntactic property that may set off the grammaticalisation of a verbal form into a quote marker and complementiser. This is, of course, only true for languages whose discourse grammar allows their (subject) pronominals to be omitted rather freely. This seems to be the case in Buru, where subjects may often be omitted in ongoing discourse if they have a contectually presupposed or known referent (Grimes, 1991: 482ff.). And it seems particularly true for the subject of the verb *fen(e)*. In the text *Gaba roi an saa* 'A notable child' (Grimes, 1991: Appendix E), sentence (39) is followed by a respons from the mother (without an element *fen(e)*), (40), and an exchange where mother and child speak in turn and the quotes are introduced by *fen* only, (41):

(39) Nak   ana-t   fene, "Ng-ina,   nau   dah.olo"
     3sPoss child-NOM say   1sVOC-mother 1sPoss bunch.head
     'Her child said, "Mother, the hand (i.e. of bananas) at the top of the stalk is for
     me"' (Grimes, 1991: 531)
(40) "Kae nam   dah.olo   moo, tu   nam   ama   nake   dah.olo"
     2s   2sPoss bunch.head NEG with 2sPoss father 3sPoss bunch.head
     "The hand at the top of the stalk isn't for you, but the hand at the top of the
     stalk is for your father"
(41) Fen, "Ng-ina,   nang   dah.dedu-k."
     say 1sVOC-mother 1sPoss bunch.repeat-k
     '(He) said, "Mother, (then) the next hand is for me."'
     Fen, "Moo, bara dah.dedu-k   tu   nam   ama"
     say NEG don't bunch.repeat with 2sPoss father
     '(She) said, "No, don't take the second hand because that's for your father."'
     Fen, "Do,   nang   dah.stifu"
     say PAUS 1sPoss bunch.middle
     '(He) said, "Well, then the middle hand is for me."'
     Fen, "Moo, bara dah.stifu   tu   nam   ama"
     say NEG don't bunch.middle with 2sPoss father
     '(She) said, "No, don't take the middle hand because that's for your father."' etc.

In this type of discourse, where the participants are both 3rd person singular and
take turns, pronominal subject marking is superfluous and may be omitted. Here, it
is the context that disambiguates, the order and the content of the quotes enables the
listeners to keep track of who is saying what.

3. Tukang Besi report constructions

The story for Tukang Besi *kua* is essentially similar to that of Buru, except for the
fact that in Tukang Besi, unlike Buru, the predicative use of *kua* has become obso-
lete. So, synchronically, there is no 'verb-like' *kua*, but only a quote marker / com-
plementiser. In (42) and (43), it follows speech act verbs such as *potae* 'say' and
wuju 'persuade', and has a quote-marking function. Note that in (42) the subject of
the verb *potae* is not marked. This example is taken from a text where the referent is
clear from the context.

(42)   Potae-mo kua   "Oho. O,   ku-wila-mo"
         say-PF   KUA yes   right 1sg-go-PF
         'She said, "Yes, I will go"' (Donohue, 1995: 489)
(43) a. No-potae -m(o) kua   "To-motindo'u na ikita"
        3R-say   -Prf KUA 1pl.R-thirsty   Nom we
        'They said "We're thirsty"'
     b. No-potae-m(o) kua   no-motindo'u na   amai
        3R-say-PF   KUA 3R-thirsty   Nom they
        'They said that they were thirsty'

In Tukang Besi, subjects are expressed by verbal prefixes (Donohue, 1995: Section 5.3.1, 5.4). Like Buru and Kambera, Tukang Besi allows subject pronominals that refer to old or background information to be dropped quite freely in casual speech (Donohue, 1995: Section 7.5.1). We return to this observation in section 4. In (44) *kua* can be interpreted as a quote marker or a conjunctive element:

(44) To-wuju-'e   kua   to-'ita-'e
     1pl.R-persuade-3Obj   KUA   1pl.R-see-3Obj
     'We persuaded her to let us see her'
     (lit.: 'We persuaded her *kua* we see her')

When *kua* follows verbs of physical (*ita* 'see') and mental (*dahani* 'know') perception, it has a complementiser function:

(45) No-'ita-'e   kua   no-kanalako te   osimpu
     3R-see-3Obj   KUA   3R-steal   Core young coconut
     'She saw that he had stolen the coconut'
(46) To-dahani   kua   no-'ita-kita   i   aba
     1pl.R-know   KUA   3R-see-1plObj Obl   before
     'We know that they saw us before'

Just like Buru *fen,* the distribution of *kua* is semantically restricted. Example (47a) shows that it cannot follow just any verb. *Kua* must follow verbs of speech or physical and mental perception. A list of these is given in (51).

(47) a. *Ku-helo'a-ke   kua   'u-manga-'e
        1s-cook-3Obj   KUA   2sg.R-eat-3Obj
        'I cooked it KUA you ate it' (Donohue, p.c.)
     b. Ku-helo'a-ke   'u-manga-'e
        1s-cook-3Obj   2sg.R-eat-3Obj
        'I cooked it (and) you ate it'

Tukang Besi clauses can be combined by simple juxtaposition, as in (47b), or by subordination. Two structures are formally marked for subordination: one is a nominalised clause (marked with a nominalising morpheme), which functions as the complement of the main verb, and the other is a controlled clause with particular verbal morphology (for details, cf. Donohue, 1995). Clause combinations with clause linkers (complementisers) such as *bara* 'lest', *ka'ano* 'in order' and *kua* in (48)–(50) are assumed to be unlike subordinations (nominalised and controlled clauses) because they lack any morpho-syntactic marker of subordination. On the other hand, they are also unlike the simple juxtaposed coordinations (where any clause combination is possible), because the semantics of the main verb determines the choice of the complementiser. We could say that the second clause is semantically, but not syntactically, dependent on the first.

(48) No-wuju-'e   bara no-wila pe'esa-no
     3R-persuade-3Obj lest 3R-go   own-3Poss
     'They persuaded him not to go on his own'
(49) No-wuju-'e   ka'ano saba'ane no-lemba-'e
     3R-persuade-3Obj in order all   3R-carry-3Obj
     'They persuaded him to carry it all'
(50) No-wuju-'e   kua   saba'ane no-lemba-'e
     3R-persuade-3Obj KUA   all   3R-carry-3Obj
     'They persuaded him that they would carry it all'

The verbs in (51) may be followed by a clause combination linked by (optional) *kua*, but they may also be followed by a nominalised or controlled clause. What is relevant here is that all the verbs allowing for *kua* as complementiser are verbs reporting speech, thought or perception (Donohue, 1995: 379, 394–395):

(51) *potae*   'say'
     pogau   'say'
     waa   'tell, command'
     wuju   'persuade'
     'elo   'call'
     'ema   'ask'
     balo   'answer'
     jandi   'promise'

     ma'eka   'fear'
     usaha   'worry' (Donohue, 1995: 492)
     dahani   'know, believe, understand' (Donohue, 1995: 522)
     roda   'remember'
     wikiri   'think' (Donohue, 1995: 512)
     putusu   'decide'
     harapu   'hope'

     ita   'see'
     rodongo   'hear'

The exact function of *kua* may vary, depending on the context. In (52a) *kua* introduces a clause that Donohue describes as a 'discourse complement', i.e. in this sentence a verbal act is implied. This clause contrasts with the 'manipulative' complement in (52b), where the persuasion could have been non-verbal as well as verbal (Donohue, 1995: 394):

(52) a. Ku-wuju-'e   kua   no-lemba'e
       1s-persuade-3Obj KUA   3R-carry-3Obj
       'I persuaded them that they carried it'

      b. Ku-wuju-'e   no-lemba-'e
        1s-persuade-3Obj 3R-carry-3Obj
        'I persuaded them to carry it'

In other words, in (52a), *kua* marks the explicit linguistic character of the act described by the verb. Now consider (53), where *kua* has a different function. In (53a) the "assertion is that the seer saw the actual fact of stealing" (Donohue, 1995: 392), contrasting with (53b), where there is no such commitment. Here, the use of *kua* implies "less direct evidential perception of the event" (ibid.), e.g. witnessing the theft because the coconut had disappeared after he had been in the room:

(53) a. No-'ita-'e   no-kanalako te   osimpu
          3R-see-3Obj 3R-steal   Core young coconut
          'She saw him steal the coconut'
      b. No-'ita-'e   kua   no-kanalako te   osimpu
          3R-see-3Obj KUA 3R-steal   Core young coconut
          'He had stolen the coconut, she saw'

The sentences in (54) have a similar contrast in 'directness'. Here the contrast is not only marked by the presence and absence of *kua*, (54a–b), but also by the fact that the second clause in (54c) is a nominalised clause:

(54) a. To-dahani   kua   no-'ita-kita   i   aba
          1pl.R-know KUA 3R-see-1plObj Obl before
          'They saw us before, we knew'
      b. To- dahani no-'ita-kita   i   aba
          1plR-know   3R-see-1plObj Obl before
          'We know they saw us before'
      c. To-dahani-'e   na 'ita-'a-no   nu   ikita
          1pl.R-know-3Obj Nom see-NL-3Poss Gen 1pl
          'We know their seeing of us'

The sentences in (54a–c) show a gradual increase in the semantic connection between the two clauses they consist of. This distinction in 'directness' is not an inherent property of *kua* itself, because it is not made when *kua* combines with speech-act verbs: in (52) the presence of *kua* only marked the linguistic character of the second clause.
    Another function of *kua* is to mark 'switch-reference' in ambiguous contexts (Donohue, 1995: 390).[12] Consider the coordination in (55). Without *kua* the subjects

---

[12] On the basis of Stirling's (1993) typological study, this use of *kua* may be better characterised as (reminiscent of) a logophoric system. Logophoric marking is usually restricted to the embedded complement clauses of a set of logocentric verbs which can be distinguished on a largely semantic basis and centrally includes verbs of reporting. Thus, dependent clauses in which logophoric pronouns are licensed are archetypically contexts of reported speech. Sometimes verbs of mental or psychological state also

of both clauses allow the interpretation of having an identical referent, with *kua* the referent must be different:

(55) a. No-roda   tabeda no-wila
        3s-remember   must   3R-go
        'She$_j$ remembered that she$_{j/k}$ had to go'
     b. No-roda   kua   tabeda   no-wila
        3s-remember   KUA   must   3R-go
        'She$_j$ remembered that she$_{*j/k}$ had to go'

To conclude, we have seen that the distribution of *kua* is semantically restricted: it follows 'report' verbs that are of the same semantic domains that we encountered in connection with Buru *fen* and Kambera *wà*. These are the same verbs that Stirling (1993) observes to be typically connected to logophoric contexts.

With illocutionary verbs, *kua* functions to mark the linguistic character of an act, with perception verbs its presence 'loosens up' the semantic connection between two adjacent clauses. In ambiguous contexts, *kua* functions as a trigger of a logophoric context; forcing the subjects of both clauses to have different referents.

These different functions of the complementiser *kua* can be explained if we assume that they are effects of the original function of *kua* as a verb reporting speech, thoughts and perceptions. It is a well-known fact that quote markers often derive from speech verbs (e.g. Lord, 1993: 151–213; Harris and Campbell, 1995: 168–172). A quote is a linguistic entity (relatively) independent of the clause by which it is introduced (Munro, 1982; see also Clark and Gerrig, 1990). And "logophoricity is often associated with the presence of a complementiser which tends to be homophonic with the verb 'say' and may originate as a reported speech opener" (Stirling, 1993: 52).

Though the predicative function of *kua* is absent in current Tukang Besi, we can find evidence that there must have been such a verb in Duri, a South Sulawesi language geographically close to Tukang Besi in South-East Sulawesi. This language has a verb *kua* 'say' which synchronically functions both as a speech verb and as a quote marker (Valkama, 1995: 77):

(56) *Duri (WMP)*
        Na-kua-an   ambeq-na   Lajanak kua "..."
        3-say-Ben   father-3Poss   Lajanak Quote
        Lajanak said to his father "..."'

So we may assume that the quote marker/complementiser *kua* in Tukang Besi has a verbal origin.

---

trigger logophoric contexts. Unlike logophoricity, switch-reference is not restricted to particular lexically governed semantic contexts (Stirling, 1993: 51–52, 256–259). Unlike 'real' logophoric languages (such as Ewe; see Duthie, 1984; Lord, 1993: 186–187; Stirling, 1993: 261), the 'logophoric pronoun' in Tukang Besi is a plain third person pronominal marker rather than a special pronoun.

Donohue (1995: 315–316) describes another element which is homophonous to *kua*: the allative[13] preposition *kua*:

(57) Maka no-waliako kua   kampo
     then   3R-return   KUA   village
     'Then they went back to the village'
(58) No-kahu-'e   kua   iaku
     3R-send-3OBJ   KUA   I
     'She sent it to me'
(59) Ku-wila   kua   ito
     1sg.R-go   KUA   there.higher
     'I'm going to the north'

Is the homophony between the preposition *kua* and the quote marker/complementiser *kua* accidental, or do the elements have a common origin? Donohue mentions no direct diachronic or synchronic evidence for a historical relation between the two. Yet, such a relation would not be implausible. Firstly, because prepositions that derive from speech verbs *have* been attested in other languages. For instance, Heine et al. (1991) mention cases of prepositions with the semantics of 'give', where 'give' is also used as a speech verb. Secondly, because Tukang Besi has other deverbal prepositions: of the five prepositions in this language, three are derived from existent verbs, while the verb *ako* 'do something for' is mentioned as a likely candidate to develop into a preposition (Donohue, 1995: 310).

My hypothesis therefore is that the homophony between the allative preposition and the quote marker/complementiser is not accidental, but suggests a common origin. In fact, in the analysis of verb grammaticalisation proposed in the next section they can be assumed to be one and the same lexical item: one of the predictions of this analysis is that an impoverished lexical element may have either a complementiser *or* a prepositional interpretation. Tukang Besi *kua* would be an example of such an item. In Section 5 I will return to this issue.

The preceding discussion can be summarised as follows. We have traced the history of the multifunctional grammatical elements *kua* in Tukang Besi and *fen* in Buru on comparative evidence and argued that in both languages these elements originally derive from verbs that report speech, thought and perception. The comparative structural analysis of *kua, fen* and *wàng(u)* provides an explanation for the semantic restrictions on the synchronic distribution of *fen* and *kua*.

Buru *fen* is less grammaticalised than Tukang Besi *kua* because it functions as verb, quote marker and complementiser, while *kua* has lost its former verbal function. On the basis of data from the related language Duri, where a speech verb *kua*

---

[13]   That is, this preposition indicates a direction rather than a destination, which would use the general locative case marker *i*:
   (i) *Ku-wil(a)*   *i*   *Waha*
      1s.R-go   Obl   Waha
      'I'm going to Waha'

is (still) in use, a verbal origin for the Tukang Besi *kua* may be assumed. The report verb in Kambera is the one least grammaticalised: it has not developed a secondary function as complementiser. In Section 4 we will offer an explanation why.

The similarity in the grammaticalisation patterns is due to the genetic relatedness of Tukang Besi, Buru and Kambera, in so far as it is has lead to a shared set of certain specific typological properties.[14] More particularly, the development of *kua/fen* is guided by the sentential and discourse syntax of the particular languages involved. The shared typological characteristics of these languages that are relevant for the grammaticalisation process of report verbs into quote markers and/or complementisers are:

(i)   They are languages which allow the pronominal morphemes that express their subject (subject-marking clitics in Buru and Kambera, subject-marking prefixes in Tukang Besi) to be omitted if these express old or background referential information (i.e. if their referents are clear from the discourse). This is a kind of 'discourse-based' pro-drop.

(ii)   They allow clause combining by simple juxtaposition.

(iii) They do not make a (morpho-)syntactic distinction between direct and indirect speech.

(iv)   They are complementiser-initial.

(v)   They have intransitive report verbs; i.e 'quote' clauses are not treated as syntactic complements.

(vi)   Their 'quote' clause follows the report verb.

The evidence for (v), the intransitive character of the report verb, was discussed in Section 1 for Kambera, and for Buru in Section 2. There is no evidence on the valency of Tukang Besi *kua* in Donohue (1995) because this element does not have a verbal function synchronically.

The last characteristic (vi) applies to Buru and Tukang Besi, but not to Kambera: though Kambera unmarked constituent order is VO, report constructions canonically use the reverse order. This is illustrated by the examples in section 1 and sentence (60), where the verb *wà* follows (rather than precedes) the quote *na-lua haromu*:

(60) Na- paní-ngga   ka   na-lua haromu   wà-na-ngga
     3sN-tell-1sD   CNJ   3sN-go tomorrow   say -3sG-1sD
     'She told me that she is leaving tomorrow'
     lit. 'She told me that "she leaves tomorrow" she said to me')[15]

Now, recall from Section 1 that Kambera *wà(ng)*, unlike Buru *fen* and Tukang Besi *kua*, has not been reanalysed as a complementiser. We can now see that the rea-

---

[14]   The idea of working out a list of shared typological characteristics of languages as the explanation for their shared historical development I owe to Miriam Butt.

[15]   The reviewer mentions as a similar case the inverted quotatives in English ('parentheticals' such as *I am ill, Mary said*). However, the Kambera case is different in that the Kambera quote *always* precedes the report verb while unmarked Kambera word order is VO.

son for this is purely structural: all three languages are complementiser-initial, so for a report verb to be re-interpreted as a complementiser it should precede the clause containing the report (or quote), as represented in (70) below. In Buru and Tukang Besi this is the case, in Kambera it is not. Even without a subject marker and with a reduced argument structure for *wà(ng)*, the verb could never be reanalysed as a complementiser, simply because the linear order of the quote and the report verb prohibits this.

In conclusion, the grammaticalisation process is steered by structural linguistic considerations. These structural properties are clusters of functionally and formally related typological properties. One particular surface syntactic property of a language may prevent a certain grammaticalisation to take place, as the Kambera case shows.

4. How report verbs become multifunctional

We have seen that the grammaticalisation of a lexical element is not an autonomous process but depends on the situation created by the particular syntax and discourse grammar of a language. In the analysis of the *kua/fen* case presented in this section we try to account for the characteristics that these items share with grammaticalised items in general:

(i)   they have reduced (inflectional) morphology;
(ii)   they have a 'bleached', more generic, semantics;
(iii) they belong to different word classes synchronically.

The idea is that loss of argument marking (for pragmatic reasons) acts as a trigger for semantic bleaching. The bleached 'verb' is now a category-neutral element that, depending on the sentential context, can be interpreted as e.g. a quote marker or a complementiser.[16] In other words, part of the grammaticalisation (here, the loss of

---

[16]   The reviewer notes that the fact that English modal verbs lack third person singular inflection (*John can /*cans do it*) and also have grammaticalised (auxiliary) functions suggests that they are a similar case. However, historical evidence shows that the modals were originally 'preterite present' forms, past tense forms that had been reinterpreted as present tense forms. Verbs of this class lacked the *–s* ending of the present (Lightfoot, 1979: 103; Harris and Campbell, 1995: 177ff.). In other words, loss of inflection marking cannot be the trigger of the grammaticalisation of English modals because there was no inflection to start with. There are other significant differences between modal verbs becoming auxiliaries on the one hand, and quote verbs becoming complementisers on the other. Firstly, though in both cases argument loss is involved, the choice of argument is different: quote verbs lost a single or external argument, while modal verbs lost an internal argument (when they "lost the ability to take direct objects", Lightfoot, 1979: 101). Secondly, the structural context of the verbs is different. In both constructions new verbs are being introduced when the old verb has become lexically impoverished, but in quote constructions the new verb (i.e. the main verb of speech or perception in the Buru/Tukang Besi examples above) precedes the original quote verb, i.e. it does not come between the quote and the quote clause, leaving open a complementiser interpretation (cf. discussion below). An lexically impoverished modal verb, however, is followed by the more specific verb, and is not in the position to get a complementiser interpretation.

argument-marking morphology) precedes the semantic bleaching of an item (elimination of an Agent/'external' argument), rather than follows it (contra Hopper and Traugott, 1993).[17]

It is a classic observation in linguistics that what is familiar and predictable is given reduced expression (e.g. Zipf, 1935). Kiparsky (1982) proposed predictability and redundancy as the two semantic principles at work in the diachronic reduction of phonological form in morphemes: "Morphological material which is predictable on the surface tends to be more susceptible to loss than morphological material which is not predictable on the surface" (1982: 67), and "There is a tendency for semantically relevant information to be retained in surface structure" (1982: 87). The former principle would be motivated by ease of production, the latter by ease of perception.

In Buru, Kambera and Tukang Besi, pronominal subjects that are clear from the discourse context can be omitted, i.e. they allow a kind of discourse-based 'pro-drop'. Especially in narrative texts, report verbs are frequently used, and there they often have subjects that are contextually predictable, so it comes as no surprise that in 'discourse pro-drop' languages the familiar and predictable subject of report verbs may be left unexpressed quite easily.

The scenario where loss of inflectional morphology triggers the grammaticalisation of speech verbs has been reported for other languages. For example, in Macedonian, the speech act verb *veli* is grammaticalising into the 'hearsay' marker *veli* (Mushin, 1997: 298) and the inflection of *veli* in evidential direct speech construction is constrained by the context to third person singulars. This has the effect of incipient loss of person/number inflection of *veli*. Mushin (1997: 296–298) shows explicitly that the loss of syntagmatic and inflectional properties of *veli* precedes the actual loss of reference to the reported speaker (the original argument of the speech act verb).

In Bislama, the English-lexified creole spoken in Vanuatu (SW Pacific), the absence of 'phonetically overt' subject marking precedes the reanalysis of verbal *se* 'say' into a complementiser (Meyerhoff, 1998: 5).

Crosslinguistically, however, there are also languages where a 'say'-verb undergoes grammaticalisation without losing the subject marker first. For example, the Georgian quote marker *metki* is formed historically from the subject-verb sequence *Me v-tkv-i* 'I said (it)' (cf. Harris and Campbell, 1995: 169 and references cited there):

---

[17] Drawing a parallel with English, the reviewer raises the question why the elimination of the theta role of a quote verb would not result in an English-type 'raising' verb such as *seem*, which is standardly analysed as lacking an external theta role with an embedded argument 'raised' to a position in the matrix clause. Assuming that this analysis is correct, the analysis of the verbs discussed in this paper suggests that there are important structural differences between English and Buru/Tukang Besi/ Kambera: (i) unlike English raising verbs, the report verbs of these languages are intransitive and lose their single argument, and (ii) unlike English, the languages discussed here are 'pronominal argument' languages which allow discourse based 'pro drop' freely, so it is to be expected that their phonologically empty arguments have different properties/restrictions than raised arguments in English. A thorough discussion of this issue is outside the scope of this paper.

(61) Ase vtkvi, ca kuxs da mic'a ikceva-**metki**
thus I.say.it sky it.thunder and earth it.tremble-QUO
'I said, "the sky thunders and the earth trembles"' or
'I said that the sky thunders and the earth trembles'
(62) Me v- tkv -i
I 1.sg.subject-say -aorist indicative
Quote marker, lit. 'I said (it)'

Because is such languages the grammaticalisation of the report verb includes a stage where the inflectional morphology is reanalysed as part of the verb or the verbal stem, we expect this type of grammaticalisation to occur in langauges that are not pronominal argument languages and do not allow 'discourse pro-drop' to occur freely. This hypothesis remains to be tested.

The initial stage in the diachronic process, where subject pronominals are omitted, was illustrated by the Buru sentences in e.g. (40). Sentence (63) is another example. *Fen* functions predicatively – not as a complementiser – but its subject is not overtly expressed.

(63) Fen, "Ng-ina, nang dah.dedu-k"
FEN 1sVoc-mother 1sPoss bunch repeat-k
'(He) said, "Mother, (then) the next hand is for me"'

In this sentence there is a mismatch between the impoverished surface form and its argument structure, because its Agent ('external', 'speaker') argument x remains unexpressed, as represented in (64). Recall that for lack of evidence to the contrary we assumed that the (reconstructed) report verbs *wà, fen* and *kua* are intransitive (see Section 1–3 above and Munro, 1982).

(64) x REPORT
| |
φ *fen*

Situations where there is a mismatch between levels of representation are inherently unstable and therefore prone to change (cf. Butt, 1997). The reason is that they violate a constraint that we call 'Semantic Transparency'. This constraint draws on the classic observation that linguistic items prefer a direct, i.e. one-to-one matching of items of different levels:

(65) *Semantic Transparency*
"Match items on different linguistic levels one-to-one"
Meaning A
|
Form X

The following configurations are examples of how this constraint can be violated. The configuration in (66d) is the one that represents (65).

(66) a.   *   A
                /\
               X.Y
          (e.g. circumfix)

b.   *   AB
           \/
           X
    (e.g. homophones)

c.   *   φ
            |
            X
    (e.g. 'meaningless' morpheme)

d.   *   A
            |
            φ
    (e.g. zero morpheme, empty
    category (e.g. *pro, PRO*))

Semantic transparency wants a mismatch like the one in (64) to be repaired. There are two ways to do this. We can add a subject pronominal at the surface form, or delete the unexpressed argument at lexical-semantic structure:

(67) a. x   REPORT
        | |
     pronominal   *fen*
     b.   REPORT
                   |
                   fen

The first strategy creates a normal clause with an overtly marked subject. The second strategy results in semantic bleaching when we delete an argument position in order to regain grammatical stability.

Note that both strategies result in a form that complies with the Transparency constraint, so there must be another reason why (eventually) the impoverished form (b) is chosen over the full form (a). This may have to do with the classic idea that simple structures are preferred over complex ones, which can be formulated as a constraint on structural simplicity:

(68) *Structural Simplicity*: "Prefer simple structures over complex ones"

Hopper and Traugott (1993) define 'semantic bleaching' as the process where lexical items lose semantic content and may become grammaticalised functional items as a result. If a verb's lexical conceptual structure is part of its 'semantic content',[18] as I will assume, 'semantic bleaching' can involve the verb's participant(s), expressed as its argument(s).[19]

---

[18] Butt (1997) contains an account of semantic bleaching where verbs lose event structure rather than argument structure.
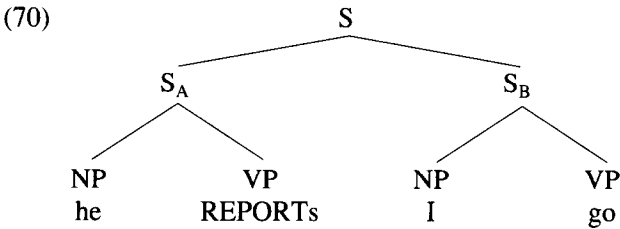
[19] The argument structure is the predicate information provided by the lexicon that is relevant for determining the syntactic behaviour of the predicate. The thematic role of the argument is irrelevant for the proposed analysis.

In the case at hand, the semantic bleaching of *kua* and *fen* is realised as loss of argument structure: the argument of the verb in (69a) is lost, resulting in the impoverished ('bleached') argument structure in (69b).
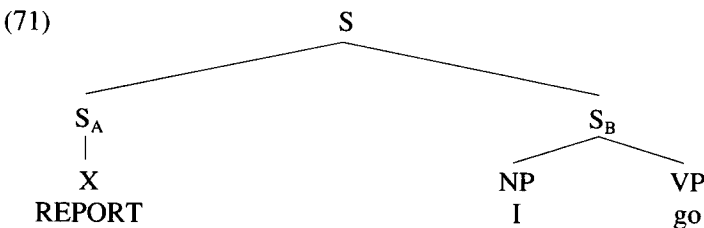
(69) a. x   REPORT
     b.   REPORT

The resulting predicate is an item with no arguments, which may be interpreted as e.g. a (result) nominal ('(there is a) report'), or a quote marker (':'), or a complementiser ('that'), depending on which syntactic context it appears in: it is interpreted as a quote marker/complementiser when it follows a verb of speech and precedes a clause, and as a complementiser when it follows perception verbs.

The following tree diagrams represent the proposed structural reinterpretation of lexically impoverished, category-neutral elements, that were quote verbs originally. In (70) a basic quotative construction (now obsolete in Buru and Tukang Besi) is represented. The quote clause S_A is introduced by a REPORT verb with subject marking.

(70)

```
                       S
            _____|_____
          S_A                      S_B
        ___|___                  ___|___
      NP       VP              NP       VP
      he      REPORTs          I        go
```
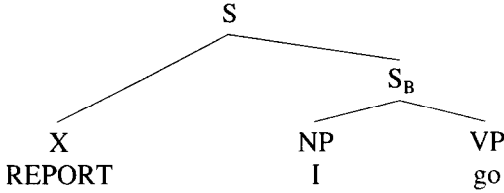
As the subject marking on the REPORT verb is omitted for pragmatic/discourse reasons, the application of the constraints on Semantic Transparency and Structural Simplicity results in the loss of the verbal argument. S_A now consists of a lexical impoverished element, a generic item REPORT only – a category-neutral item without derivational or inflectional morphology.
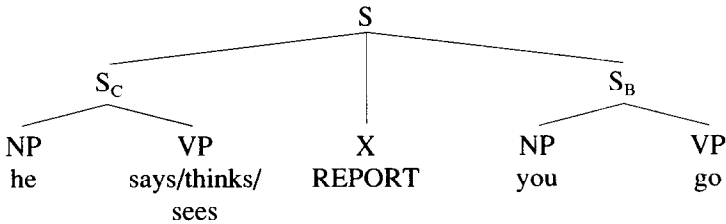
(71)

```
                       S
          _____|_____
        S_A                          S_B
         |                         ___|___
         X                        NP      VP
      REPORT                      I       go
```

As the clausal node S_A has no daughters other than X, there is no overt evidence for this node. Unless there is intonational evidence to identify the node S_A as a separate clause, as e.g. in (41), it becomes uninterpretable. There is no longer evidence to assume that the REPORT predicate constitutes a separate clause, and then Structural Simplicity ensures that the S_A node disappears:

(72)
```
                        S
              _____/ _____
             /                   \
            /                    S_B
           /                    /   \
          X                   NP     VP
        REPORT                 I      go
```

Because the old bleached report verb lost its argument, its predicative potential also got lost. New verbs are now introduced to replace the old one.[20] The new verbs are not (yet?) semantically bleached, take subject marking and make up the predicate of clause S_C : 'He says (that) you should go'/'He thinks that you go'/'He sees you go'):

(73)
```
                           S
        _____/|_____
       /                   |                   \
      S_C                   |                   S_B
     /   \                  |                  /   \
   NP     VP                X                NP     VP
   he   says/thinks/      REPORT            you     go
          sees
```

The REPORT item is now a category-neutral element in between two clauses, which makes the complementiser reading available. Recall that the complementiser reading of *kua/fen* was especially clear when they are preceded by a perception verb, followed by a clause, and not separated from that clause by an intonational break. If the main verb is a verb reporting speech, *kua/fen* may be interpreted as a quote marker. In this way, the context of the REPORT element determines its interpretation. The synchronic multifunctional characteristics of the grammaticalised item are a consequence of the interaction between its lexically impoverished structure and surface syntax. Also, the 'category change' from verb to quote marker or complementiser is only apparent – there is no actual change in the lexical representation of the underspecified category-neutral element X. It may be that the 'category change' or 'reanalysis of category labels' (Harris and Campbell, 1995: 63) that is often part of a grammaticalisation process can be analysed in a similar way – as the result of the interpretation of a category neutral element by surface syntax.

---

[20] Other analyses of grammaticalisation of speech verbs (e.g. Crowley, 1989) assume the speech verb grammaticalises when it appears in combination with a semantically more specific verb, i.e. in the context [V-specific V-REPORT]. In such an analysis, the stage represented by sentence (63) would not be expected. In the present analysis, however, this sentence represents the initial stage of the grammaticalisation of *fen* (and *kua*). Consequently, the bleaching of report verbs is assumed to *precede* the introduction of the semantically specific verbs (rather than be a result of it).

5. Future research and conclusions

The proposal to view grammaticalised verbs as forms with a reduced argument structure can be extended to other areas of verb-grammaticalisation as well. In what follows I point out some general consequences of the analysis and potential directions of future research.

One instance where 'semantic bleaching' can be assumed to involve loss of argument structure is the common case where the second verb in a serial verb construction also develops a prepositional function. The change involved is similar to the one we have discussed. Let us make the assumption that both prepositions and complementisers have one argument, while transitive verbs have two:[21]

(74) Transitive verb: x   X   y
      Preposition/complementiser: X   y

Thus, prepositions and complementisers have the same argument structure, but their subcategorisation properties are different: canonically, complementisers take a clausal complement with propositional semantics, while prepositions take a nominal complement referring to an entity.

For a verb in a serial verb construction to become prepositional, the initial stage is when it loses its own subject marking. This is an effect of the verb sharing its subject with the other verb in the serial verb construction – subject sharing is one of the distinctions between a serial verb and a biclausal construction. For example, consider the following Kambera serial verbs, where the second verb is instrumental *wàngu* 'use' (Klamer, 1998: 288, 290):[22]

(75) taku wàngu[23]
      scoop use   'scoop X using Y'
      meti wàngu
      die use   'kill X using/because of Y'
      riki wàngu
      laugh use   'laugh about Y'
      hayidi wàngu
      play.games use   'play games on Y'

The subject sharing of the verbs in a serial verb such as *riki wàngu* 'laugh about Y' can be represented as follows:

---

[21] But recall that Buru/Tukang Besi *kua/fen* are assumed to derive from *intransitive* verbs.

[22] For a full discussion of the properties of Kambera serial verbs, including the instrumental ones, I refer to Klamer (1998: 284–295).

[23] Note the near homophony between the quote verb *wàng* and the instrumental verb *wàngu* 'use'. I do not know whether this is accidental or points to a common source.

(76) subject   ['laugh'   φ   'use'] object
       | | | | |
      [x   'laugh']   [x   'use'   y]

Here there is a mismatch between conceptual structure where there are two *x* (agent, external) arguments (and one *y* or non-agent, internal), and surface structure where the agent arguments are expressed only once as the shared subject. This mismatch is identical to the mismatch that triggered the change in the report verbs. The repair strategy is the same as well: the second verb loses the unexpressed argument and becomes lexically impoverished, making room for a prepositional interpretation and resulting, for example, in the possibility to project as a separate prepositional phrase.[24] This is illustrated for Kambera in (77). In (77a) the verbal root *wà* is still part of the serial verb construction, but in (77b) it heads a preposed PP:

(77) Ku- **taku** **wà-nya**ⱼ uhu   [na   huru-mu   nyumu]ⱼ
     1sN- scoop use-3sD   rice ART   spoon-2sG you
     'I scoop rice using/with your spoon'
     [**Wà-nya**ⱼ [na   huru-mu   nyumu]ⱼ]ₚₚ ku-   **taku** uhu
     use-3sD   ART   spoon-2sG you   1sN- scoop rice
     'With/using your spoon I scoop rice'

Why some verbs become reanalysed as complementisers while others become prepositions is configurationally determined. Report verbs are canonically followed by the clause which contains the report (or 'quote'), i.e. their context allows them to be interpretated as complementisers, while the second verb in a transitive serial verb construction is followed by a nominal rather than a clausal constituent, which, of course, allows for a prepositional interpretation. Note, however, that the Kambera illustration also shows that the 'following' constituent is not necessarily the NP which follows the verb (here *uhu* 'rice') but may be the NP that is crossreferenced by a clitic on the verb (*-nya* refers to *na huru-mu nyumu* 'your spoon') .

In this analysis, one and the same underspecified lexical-conceptual structure can in principle have a complementiser *or* prepositional interpretation, depending on what follows it. In Section 3 we saw that Tukang Besi *kua* may be an example of such an item. *Kua* is interpreted as a quote marking/complementiser when it precedes a clause, and as a preposition when it precedes a nominal constituent.[25]

---

[24] Needless to say, this is not meant to say that all prepositions stem from verbs – prepositions may also have adjectival (e.g. Maling, 1983) or nominal ancestors. Denominal prepositions commonly stem from relational nouns ('side', 'back', etc. in e.g. the Indo-Aryan languages, Miriam Butt, p.c.), and as such can be argued to have an argument position in their conceptual structure, just like adjectives.

[25] A similar case is French *de*, functioning as a complementiser in infinitival contexts and as a preposition elsewhere. Also, in Maori (New Zealand), the item *ki* functions both as a dative-allative particle and as a complementiser (Hopper and Traugott 1993):
  (i) E   hoki   ana   au **ki** te   kaainga
       PRES return PROG   I   KI   the   village

A complication in this analysis is that we assumed that the category neutral element *kua* is intransitive, while a standard prepositional reading would imply *kua* to have a complement. This can be solved by assuming that the original report verb was transitive after all, but in doing so we would deny the complete lack of evidence for the transitive status of *kua*.

Another option is to assume that lexical elements can *acquire* (as well as lose) argument structure, which is what I propose to be the case here. It has often been observed that multifunctional items may semantically drift apart in the course of time; each of them acquiring a new, specific interpretation and different subcategorisation properties. In terms of argument structure, then, underspecified and lexically poor elements may *get* new arguments as well as lose them. In both cases, this is the result of frequently occurring in appropriate surface contexts. Thus, the surface grammar triggers a change in argument structure of the category neutral element *kua*, now adding an argument (cf. Vincent's (1998) analysis of the development of Latin particles into prepositions, which also assumes a stage of argument aquiring). In this view, the lexicon is not a list of static items, but lexical argument structure is a dynamic identity, formed through interaction with syntax.

The English adjective *like* is another example where a (probably) underspecified item allows a variable interpretation (cf. Maling, 1983).[26] In (78) *like* is presented in a number of different contexts, and it is clear that (stripped from its adjectival inflections; Maling, 1983: 277) this adjective also allows a C or P (or Adv) interpretation. The distinction between A, C and P (and Adv), if it exists, is gradual and entirely determined by the syntactic context of *like*:

(78) Cry like a baby   like = A or P?
     He ran like crazy   like = A, P or C?
     It looks like he will win   like = C
     I wouldn't mind, it's just like I prefer not to   like = C
     It goes like "beng"   like = C, P, or Adv?

The above does not imply that 'anything goes' in terms of argument structure change, and that we can randomly postulate acquired or lost arguments in lexical items to 'explain' historical change or grammaticalisation. We need to be specific about what *exactly* determines the argument loss or acquisition, and define when a variable interpretation may or may not result.

For the report verbs discussed in this paper I first proposed two general linguistic constraints on the form-content mapping[27] in linguistic signs. These principles (Semantic Transparency and Structural Simplicity) complement the (discourse-

---

[26] Despite the homophony, verbal *like* (OE *lician*, cf. Allen, 1995) and adjectival *like* (OE *gelic*; Maling, 1983) are not cognates.

[27] These constraints prefer a simple mapping of content unto form to a complex mapping, and are clearly related to the Structural markedness constraints that have been proposed in recent work in Optimality Theory (Prince and Smolensky, 1983; cf. Bresnan, 1998).

based) loss of pronominal argument marking on a verb by the loss of the verb's argument structure.

Then I argued that the interpretation of an underspecified lexical item depends on the specific syntactic context in which it appears, and I defined these contexts in terms of phrase structure. The multifunctionality of a grammaticalising verb is thus an expected property, not a problematic one; and so-called 'V to C' grammaticalisation involves contextually determined variable interpretation rather than lexical category change.

Thirdly, I suggested a very specific list of typological properties of these languages that make them liable for this particular type of verb grammaticalisation. And finally, I explained that, despite of the many similarities between Kambera and Buru and Tukang Besi, Kambera *wà* did not grammaticalise because of the different (surface) constituent order of Kambera quotes. Genetic relatedness is thus not a guarantee for similar developments.

References

Adelaar, Willem F.H., 1990. The role of quotations in Andean discourse. In: Harm Pinkster and Inge Genee (eds.), Unity in diversity, 1–12. Dordrecht: Foris.

Allen, Cynthia L., 1995. Case marking and reanalysis. Oxford: Clarendon.

Berg, René van den, 1985. A grammar of the Muna language. Dordrecht: Foris.

Bresnan, Joan, 1998. Optimal syntax. To appear in: Joost Dekkers, Frank van der Leeuw and Jeroen van de Weijer (eds.), Optimality theory: Phonology, syntax and acquisition. Oxford: Oxford University Press.

Butt, Miriam, 1997. Interfaces as locus of historical change. Proceedings of the LFG97 Conference. Stanford, CA: CSLI Publications.

Clark, Herbert H. and Richard J. Gerrig, 1990. Quotations as demonstrations. Language 66(4), 764–805.

Coulmas, Florian (ed.), 1986. Direct and indirect speech. Berlin: Mouton De Gruyter.

Crowley, Terry, 1989. Say, c'est, and subordinate constructions in Melanesian pidgin. Journal of Pidgin and Creole Languages 4(2), 185–210.

De Roeck, Marijke, 1994. A functional typology of speech reports. In: Elisabeth Engberg-Pedersen, Lisbeth Falster Jakobsen and Lone Schack Rasmussen (eds.), Function and expression in Functional Grammar, 331–351. Berlin: Mouton de Gruyter.

Donohue, Mark, 1995. The Tukang Besi language of Southeast Sulawesi, Indonesia. Ph.D. dissertation, Australian National University.

Duthie, A.S., 1984. Displaying the semantic structure of an Ewe text. Journal of West African Languages 14(1), 57–79.

Engelenhoven, Aone van, 1995. A description of the Leti language (as spoken in Tutukei). Ph.D. dissertation, Leiden University.

Grimes, Charles E., 1991. The Buru language of Eastern Indonesia. Ph.D. dissertation, Australian National University.

Harris, Alice C. and Lyle Campbell, 1995. Historical syntax in cross-linguistic perspective. Cambridge: Cambridge University Press.

Haspelmath, Martin, 1998. Does grammaticalization need reanalysis? Studies in Language 22(2), 315–351.

Heine, Bernd and Mechtild Reh, 1984. Grammaticalization and reanalysis in African languages. Hamburg: Buske.

Heine, Bernd, Ulrike Claudi and Friederike Hünnemeyer, 1991. Grammaticalization: A conceptual framework. Chicago, IL: University of Chicago Press.

Hopper, Paul and Elisabeth Traugott, 1993. Grammaticalization. Cambridge: Cambridge University Press.

Kiparsky, Paul and Carol Kiparsky, 1970. Fact. In: Danny D. Steinberg and Leon A. Jakobovits (eds.), Semantics, 345–369. Cambridge: Cambridge University Press.

Kiparsky, Paul, 1982. Explanation in phonology. Dordrecht: Foris.

Klamer, Marian, 1994. Kambera: A language of Eastern Indonesia. Ph.D. dissertation, Vrije Universiteit Amsterdam.

Klamer, Marian, 1996. Kambera has no passive. In: Marian Klamer, (ed), Voice in Austronesian [NUSA 39], 12–30. Jakarta: Universitas Atma Jaya.

Klamer, Marian, 1998. A grammar of Kambera. Berlin: Mouton de Gruyter.

Klamer, Marian, 1999. Report constructions in Kambera (Austronesian). To appear in: Tom Güldemann and Manfred von Roncador (eds.), Reported discourse: A meeting ground for different linguistic domains. Amsterdam: Benjamins.

Klinken, Catharina van, 1997. A grammar of the Fehan dialect of Tetun. Ph.D. dissertation, Australian National University.

Lightfoot, David, 1979. Principles of diachronic syntax. Cambridge: Cambridge University Press.

Lord, Carol, 1993. Historical change in serial verb constructions. Amsterdam: Benjamins.

Maling, Joan, 1983. Transitive adjectives: A case of categorial reanalysis In: Frank Heny and Barry Richards, Linguistic categories: Auxiliaries and related puzzles. Dordrecht: Reidel.

Meyerhoff, Miriam, 1998. All the same? The emergence of complementisers in Bislama. Ms. University of Hawaï at Manoa.

Mushin, Ilana, 1997. Direct speech and evidentiality in Macedonian. In: Kora Singer, Randall Eggert and Gregory Anderson (eds.), Chicago Linguistic Society 33, Papers from the main session, 287–300.

Munro, Pamela, 1982. On the transitivigy of 'say' verbs. In: Paul J. Hopper and Sandra M. Thompson (eds.), Syntax and semantics: Studies in transitivity, 301–318. New York: Academic Press.

Noonan, Michael, 1985. Complementation. In: Timothy Shopen (ed.), Language typology and syntactic description. Volume II, 42–140. Cambridge: Cambridge University Press.

Ohlander, S., 1976. Phonology, meaning, morphology: On the role of semantic and morphological criteria in phonological analysis. Göteborg: Acta Universitatis Gothoburgensis.

Prince, Allan and Paul Smolensky, 1993. Optimality theory: Constraint interaction in generative grammar. RuCCS Technical report #2. Piscateway, NJ: Rutgers University Center for Cognitive Science.

Reesink, Ger P., 1993. 'Inner speech' in Papuan languages. Language and Linguistics in Melanesia 24, 217–225.

Roberts, Ian, 1985. Agreement patterns and the development of English modal auxiliaries. Natural Language and Linguistic Theory 3, 21–58.

Smolensky, Paul, 1996. The initial state and 'richness of the base' in Optimality Theory. Technical Report JHU-CogSci-96-4, Department of Cognitive Science, Johns Hopkins University, Baltimore, MD.

Stirling, Lesley, 1993. Switch-reference and discourse representation. Cambridge: Cambridge University Press.

Thomas, E., 1978. A grammatical description of the Engenni language. Dallas, TX: Summer Institute of Linguistics.

Traugott, Elisabeth and Bernd Heine, 1991. Approaches to grammaticalization. Amsterdam: Benjamins.

Tryon, Darrell T., 1995. Comparative Austronesian dictionary: An introduction to Austronesian studies. Berlin: Mouton de Gruyter.

Valkama, Kari, 1995. Person marking in Duri. In: René van den Berg (ed.), Studies in Sulawesi linguistics Part IV [NUSA 37], 47–95. Jakarta: Universitas Atma Jaya.

Vincent, Nigel, 1998. The evolution of c-structure: Prepositions and PPs from Indo-European to Romance. Ms. University of Manchester.

Vries, Lourens de, 1990. Some remarks on direct quotation in Kombai. In: Harm Pinkster and Inge Genee (eds.), Unity in discourse diversity, 291–308. Dordrecht: Foris.

Vygotsky, L.S., 1962. Thought and language. Cambridge, MA: The MIT Press.

Walker, Alan T., 1982. A grammar of Savu. Pacific Linguistics. Canberra: Dept. of Linguistics RSPacS, Australian National University.

Woollams, Geoff, 1996. A grammar of Karo Batak, Sumatra. [Pacific Linguistics C-130.] Canberra: Dept. of Linguistics RSPacS, Australian National University.

Zipf, George, 1935. The psychobiology of language. Boston, MA: Hoghton Mifflin.