J. Linguistics **47** (2011), 587–630. © Cambridge University Press 2011
doi:10.1017/S0022226711000016 First published online 10 March 2011

On the grammaticalization of personal pronouns[1]

BERND HEINE

Universität zu Köln

KYUNG-AN SONG

Chonnam National University

(Received 7 April 2010; revised 12 November 2010)

Unlike most other grammatical domains, that of personal pronouns is clearly underresearched in works on grammaticalization. One reason can be seen in the fact that personal pronouns differ in their diachronic behavior from most other grammatical categories to the extent that they present a challenge to grammaticalization theory. In the present paper it is argued that in order to account for this behavior, an extended understanding of grammaticalization is needed.

1. Introduction

Grammaticalization theory has been used to describe and explain a wide range of linguistic phenomena. While the development of person agreement markers has found some attention in this work (see Siewierska 2004 for a summarizing discussion), the question of how personal pronouns arise has essentially been ignored so far, especially for the following reason: Markers for personal deixis belong to the most conservative parts of grammar, that is, they are diachronically fairly stable, as is suggested by the fact that in many cases they can be traced back etymologically to or even beyond the earliest stages of reconstructable language history. This is why so far not much progress has been made in unraveling the development of personal

[1] We wish to express our gratitude to a number of colleagues who have been of help in writing this paper, in particular to John Haiman, Christa König, Tania Kuteva, Heiko Narrog, Fritz Newmeyer, as well as two anonymous *JL* referees. We also wish to express our gratitude to the Korean Ministry of Education, Science and Technology for generously having sponsored the research leading to this paper within its World Class University Program.

The following abbreviations are used in the paper: 1, 2, 3 = first, second, third person; ABS = absolute pronoun; ACC = accusative; CL = classifier; CONN = connective; COP = copula; DAT = dative; DEF = definite; DU = dual; EMPH = emphasis; END = ending; EX = exclusive; F = feminine; FUT = future; GNS = grammaticalization in a narrow sense; GWS = grammaticalization in a wide sense; H = honorific; HAB = habitual; HUM = human; IN = inclusive; INT = intensifier; INTR = interrogative particle; M = masculine; N = neuter; NEG = negation; NOM = nominative; O = object; OBJ = object; PAST = past; PFV = perfective; PL = plural; PRES = present; REFL = reflexive; SG = singular; SUB = subject; TAM = tense-aspect-modality; TNS = tense; TOP = topic; TR = trial.

pronouns; the number of cases where it has been possible to determine their origin is severely limited. Still, some data do exist and allow us to reconstruct the main contours of this development.

The present paper is organized as follows. In Section 2 we look at the diachronic sources of each of the main categories of personal deixis based on a crosslinguistic perspective. The more general implications of our survey are the subject matter of Section 3, where we will be concerned with the question of how, or to what extent, the development of personal pronouns is covered by grammaticalization theory. In the final Section 4 then we will highlight some general findings presented and draw attention to areas where more research is required.

1.1 *On personal pronouns*

Sets of personal pronouns can be found in most languages of the world, but they are perhaps most diversified and complex in societies characterized by pronounced forms of hierarchical social organization and status. In such societies, distinctions in personal reference and address are likely to thrive, paradigm cases being found in Southeast and East Asia. Languages such as Thai, Burmese, Khmer, Vietnamese, Korean, and Japanese dispose of complex systems of personal pronouns based on distinctions of honorification (see e.g. Forchheimer 1953, Cooke 1968, Head 1978, Haase 1994, Cysouw 2003, Siewierska 2004, Heine & Song 2010), and a considerable part of the data to be discussed below are taken from languages spoken by these societies.

Our interest here is mainly with independent personal pronouns (henceforth in short: 'personal pronouns'), that is, with items having the following properties: (a) they are words having their own prosody, (b) their primary or only function is to express distinctions in personal deixis, (c) they lack specific semantic content, (d) they resemble noun phrases in their positional possibilities but do not normally take modifiers, (e) they form a closed class (Sugamoto 1989; Heath 2004: 1002; Helmbrecht 2004).

Languages are generally assumed to have personal pronouns.[2] We will treat them as a sub-class of person markers, that is, linguistic elements that are shifters which are specialized for this function and are used for the expression of personal deixis (cf. Cysouw 2003: 5).[3] Person markers include a

[2] Two languages are reported to lack personal pronouns, namely the Keresan language Acoma of New Mexico and the Chapacura-Wanhan language Wari' of Brazil (Heath 2004: 999).

[3] Person markers are linguistic elements that are shifters which are specialized for this function and are used for reference to speech act participants (Cysouw 2003: 5).

wide range of grammatical elements other than personal pronouns, such as bound markers, agreement markers, etc.

The boundaries between personal pronouns and other kinds of person markers are fluid. Siewierska (2004: xv) rightly points out that 'different instantiations of the category of person are best viewed as defining both a diachronic and a synchronic cline in regard to their formal and functional properties'. Unfortunately, there is so far no diachronic reconstruction of this cline based on crosslinguistic data. Whether or not a given element conforms in every respect to our definition is not a matter of major concern in this paper.

The grammatical status of personal pronouns has been discussed controversially.[4] We will assume that they are functional categories rather than, say, a subclass of nouns or other categories,[5] and in doing so we are relying on the parameters of grammaticalization to be discussed in Section 1.2:

(i) Unlike nouns, personal pronouns have a schematic meaning that can be described fairly exhaustively in terms of a few elementary conceptual distinctions, most of all relating to personal deixis and number (desemanticization).
(ii) Personal pronouns have a more restricted categorial potential than lexical categories, frequently lacking e.g. the ability to take modifiers or inflectional and derivational affixes (decategorialization).
(iii) Personal pronouns are as a rule shorter than nouns and verbs (erosion).

A number of generalizations have been proposed on person markers, in particular the following:

(1) *Some generalizations on personal pronouns*
  (a) If a language has gender distinctions in first person pronouns, it always has gender distinctions in the second or third person, or in both. (Greenberg 1963b, Universal 44)
  (b) If a language has a gender distinction in the first and second person, it also has one in the third person.
  (c) A language never has more gender categories in non-singular numbers than in the singular. (Greenberg 1963b, Universal 37)
  (d) If a singular and corresponding plural form are suppletive then this is more likely to apply to the first than to the second or third person. (Forchheimer 1953: 65; Heath 2004: 1008)

[4] We are grateful to Fritz Newmeyer for having drawn our attention to this issue.

[5] We are ignoring here views based on specific theoretical assumptions that lead the authors concerned to distinguish between different levels of analysis. For example, on the basis of syntactic evidence, Postal (1969) concluded that English does not have personal pronouns. On this view, 'surface structure' forms such as *he*, *she*, or *we* are in their 'deep structure' articles similar to *the* and *a*.

(e) The majority of the languages of the world show person marking on the verb.[6]

There are two contrasting positions on the typological status of person markers (see Cysouw 2003: 28ff. for details). At one end there are those who maintain that 'I' and 'you' are universal in some sense, and that grammaticalized expressions for them are to be expected in any language. This is a view that surfaced in crosslinguistic surveys of the mid-20th century. Thus, Forchheimer (1953: 39) maintained that '[t]he distinction of speaker, addressed, and neither speaker nor addressed is universally found' and Greenberg's (1963b: 90) Universal 42 also assumes the presence of three persons and two numbers to be characteristic of pronominal categories in all languages. At the other end there are those who have expressed doubts on whether these deictic concepts are really universal (Mühlhäusler & Harré 1990; Cysouw 2003: 13), and Cysouw (2003: 27) concludes that every language can mark person in some way, 'but whether every language has a category of person remains unproven'.

1.2 *Grammaticalization*

Grammaticalization is defined as the development from lexical to grammatical forms, and from grammatical to even more grammatical forms.[7] Since the development of grammatical forms is not independent of the constructions to which they belong, the study of grammaticalization is in the same way concerned with constructions, and with even larger discourse segments (see Heine, Claudi & Hünnemeyer 1991; Traugott & Heine 1991a, b; Bybee, Perkins & Pagliuca 1994; Lehmann 1995; Kuteva 2001; Heine & Kuteva 2002; Hopper & Traugott 2003 for details). In accordance with this definition, grammaticalization theory is concerned with the genesis and/or development of grammatical forms. Its primary goal is to describe how grammatical forms and constructions arise and develop through space and time, and to explain why they are structured the way they are. One main motivation for grammaticalization consists in using linguistic forms for meanings that are concrete, easily accessible, and/or clearly delineated to also express less concrete, less easily accessible and less clearly delineated meaning contents. To this end, lexical or less grammaticalized linguistic expressions are pressed into service for the expression of more grammatical functions; we will return to this issue in Section 3.

---

[6] According to the sample of 378 languages of Siewierska (2004: 414), only 21.7% (82) of all languages show no person marking of any argument, while 51.1% (193) of the languages mark person on both A (= subject of a transitive clause) and the P (= object of a transitive clause) arguments.

[7] For a fairly comprehensive list of definitions that have been proposed for grammaticalization, see Campbell & Janda (2001).

In order to identify processes of grammaticalization, a wide range of criteria have been proposed (see e.g. Lehmann 1985, 1995; Hopper 1991; Bybee et al. 1994; Heine & Kuteva 2002; Hopper & Traugott 2003). A number of notions have been proposed, such as syntacticization, morphologization, obligatorification, subjectification, etc., and several alternative frameworks offer convenient tools for describing grammaticalization phenomena. We believe, however, that they are either too restrictive, thereby excluding a number of such phenomena (e.g. Lehmann 1985, 1995)[8] or too general (Hopper 1991) to isolate the range of processes that we consider to be cases of grammaticalization.[9] In our model it is the four parameters listed in (2) which, as we argue, take care of the relevant criteria that have been proposed in other frameworks. Henceforth we will rely on these parameters, using them as a tool for identifying instances of grammaticalization.

(2) *Parameters of grammaticalization*
  (a) Extension, i.e. when linguistic expressions are extended to new contexts that invite the rise of grammatical functions (context-induced reinterpretation).
  (b) Desemanticization (or 'semantic bleaching'), i.e. loss (or generalization) in meaning content.
  (c) Decategorialization, i.e. loss in morphosyntactic properties characteristic of lexical or other less grammaticalized forms.
  (d) Erosion ('phonetic reduction'), i.e. loss in phonetic substance.[10]

Each of these parameters concerns a different aspect of language structure or language use; (2a) is pragmatic in nature, (2b) relates to semantics, (2c) to morphosyntax, and (2d) to phonetics. Except for (2a), these parameters all involve loss in properties. But the process cannot be reduced to one of structural degeneration. There are also gains. In the same way as linguistic items undergoing grammaticalization lose in semantic, morphosyntactic, and

[8] Lehmann (1985, 1995[1982]), for instance, proposes criteria such as condensation (decrease in structural scope), coalescence (increase in bondedness), and fixation (decrease in syntactic variability). While they do in fact apply to a number of processes, there are other grammaticalization processes, relating e.g. to the evolution of adverbial material, tense–aspect marking, and discourse structure, where these criteria may not apply (see e.g. Barth-Weingarten & Couper-Kuhlen 2002: 357; Günthner & Mutz 2004; Hengeveld in press; Norde in press). Lehmann's criteria coalescence and fixation are both manifestations of our parameter of decategorialization (see (3d, e) below). A really discriminating criterion in Lehmann's catalogue is integrity (relating to semantic and phonetic attrition). As we will see below, however, semantic attrition (i.e. desemanticization; see below) and phonetic attrition (i.e. erosion) should be strictly separated because the former is a central feature of grammaticalization processes whereas the latter does not constitute a requirement.

[9] With the exception of decategorialization, a term that the first-named author suggested to Paul Hopper in 1989, all principles in Hopper (1991) also apply to some other phenomena of lexical and grammatical change.

[10] As rightly pointed out to us by an anonymous *JL* referee, phonetic reduction can also be due to regular sound change.

phonetic substance, they also gain in properties characteristic of their uses in new contexts – to the extent that in some cases their meaning and syntactic functions may show little resemblance to their original use.

The ordering of these parameters reflects the diachronic sequence in which they typically apply. Grammaticalization tends to start out with extension, which triggers desemanticization, and subsequently decategorialization and erosion. Erosion is the last parameter to come in when grammaticalization takes place, and in a number of the examples to be presented below it is not (or not yet) involved. Paradigm instances of grammaticalization involve all four paramaters but, as we will see in Section 2.3.2.2, there are also cases where not all of the parameters play a role. We will now look at each of these parameters in turn.

1.2.1 *Extension*

Of all the parameters, extension is the most complex one, it has a sociolinguistic, a text-pragmatic, and a semantic component. The sociolinguistic component concerns the fact that grammaticalization starts with innovation (or activation) as an individual act, whereby some speaker (or a small group of speakers) proposes a new use for an existing form or construction, which is subsequently adopted by other speakers, ideally diffusing throughout an entire speech community (= propagation; see e.g. Croft 2000: 4–5). The text-pragmatic component involves the extension from a usual context to a new context or set of contexts, and the gradual spread to more general paradigms of contexts. The semantic component finally leads from an existing meaning to another meaning that is evoked or supported by the new context; we will return to this parameter in Section 3.

1.2.2 *Desemanticization*

Desemanticization is frequently triggered by metaphoric processes (Lakoff & Johnson 1980, Lakoff 1987, Heine et al. 1991). For example, a paradigm case of grammaticalization involves a process whereby body part terms ('back', 'breast', etc.) are reinterpreted as locative adpositions ('behind', 'in front', respectively) in specific contexts. Via metaphorical transfer, concepts from the domain of physical objects (body parts) are used as vehicles to express concepts of the domain of spatial orientation (extension), while desemanticization has the effect that the concrete meaning of the body parts is bleached out, being reduced, or giving way, to some spatial schema.

1.2.3 *Decategorialization*

Once a linguistic expression has been desemanticized, e.g. from a lexical to a grammatical meaning, it tends to lose morphological and syntactic

properties characterizing its earlier use but being no longer relevant to its new use. Decategorialization entails in particular the changes listed in (3).

(3) *Salient properties of decategorialization*
  (a) Loss of the ability to be inflected.
  (b) Loss of the ability to take on derivational morphology.
  (c) Loss of ability to take modifiers.
  (d) Loss of independence as an autonomous form, increasing dependence on some other form.
  (e) Loss of syntactic freedom, e.g. of the ability to be moved around in the sentence in ways that are characteristic of the non-grammaticalized source item.
  (f) Loss of ability to be referred to anaphorically.
  (g) Loss of members belonging to the same grammatical class.

In accordance with this list, nouns undergoing decategorialization tend to lose morphological distinctions of number, gender, case, definiteness, the ability to combine with adjectives, determiners, etc., to be headed by adpositions, they lose the syntactic freedom of lexical nouns, and the ability to act as referential units of discourse.

Verbs undergoing decategorialization tend to lose their ability to inflect for tense, aspect, negation, etc., to be morphologically derived, to be modified by adverbs, to take auxiliaries, to be moved around in the sentence like lexical verbs, to conjoin with other verbs, to function as predicates, and to be referred to by e.g. pro-verbs. Finally, they change from open-class items to closed-class items.

In more general terms, decategorialization tends to be accompanied by a gradual loss of morphological and syntactic independence of the linguistic item undergoing grammaticalization, typically proceeding along the scale described in (4).

(4) Free form > clitic > affix

1.2.4 *Erosion*

As a result of grammaticalization, a linguistic expression tends to lose parts of its morphophonological substance. We observed above that this parameter is usually the last to apply, and it is not a requirement for grammaticalization to happen. Erosion can be morphological or phonetic. In the former case it leads to the loss of entire morphological elements, and in the latter to the loss of phonetic properties (see Heine & Reh 1984). Phonetic erosion involves any of the processes listed in (5), or some combination thereof.

(5) *Kinds of erosion*
  (a) Loss of phonetic features or segments, including loss of full syllables
  (b) Loss of suprasegmental properties, such as stress, tonal distinctions, or intonation
  (c) Loss of phonetic autonomy and adaptation to adjacent phonetic units

2. Sources

The present section is concerned with grammatical evolution. We will try to trace the origin of personal pronouns back to their genesis and to account for their structure with reference to their history. To this end, we will deal with each deictic category in turn, where the three categories are first, second, and third person. For good reasons, third person pronouns are distinguished in many works from the other two categories of personal deixis, most of all because they do not refer to speech participants and crosslinguistically tend to differ in their morphosyntactic behavior from first and second person pronouns. However, we agree with Siewierska (2004: 8) in arguing that there is no clear advantage in excluding third person pronouns from the category of person and, as we hope to demonstrate below, this is also borne out by the study of their evolution.

In describing the evolution of personal deixis we will rely on the categories proposed in Heine & Song (2010; see also Helmbrecht 2004, 2005a, b), which are hypothesized to provide the main conceptual sources for personal pronouns, namely those listed in (6).

(6) *The main conceptual sources of personal pronouns*
  (a) Nominal concepts
  (b) Spatial deixis
  (c) Intensifiers
  (d) Plurification
  (e) Shift in deixis

The primary concern in this paper is with the genesis of new personal pronouns. Not all instances of such a process that have been reported, however, are in fact of this kind. For example, Shibatani (1990: 31) says that the personal pronouns of the Ainu language of Japan are derived from any of several existential verbs meaning 'to exist'. For example, the (colloquial) first person pronoun *kuani* of Ainu is analyzable as consisting of three components: *ku-* (transitive affix of first person singular), *an* 'exist', and the nominalizing suffix *-i*. We will ignore cases such as Ainu in this paper, however, because they do not lead to new categories of personal deixis; rather, they are concerned with the modification of already existing categories. Thus, prior to the rise of *kuani* 'I' there was already an etymologically

related first person singular marker **ku-* in Ainu which contributed to the form of the new person marker.

Processes such as the one illustrated with this Ainu example, which do not affect the category of personal deixis concerned, also fall within the scope of grammaticalization theory, being referred to as renewal or 'strengthening'. But they are not responsible for genuinely new personal pronouns, that is, for the genesis of personal pronouns from forms that previously either were not personal pronouns or were pronouns denoting a different category of personal deixis (see Section 2.2.1 for examples of the latter process). Pronoun renewal is quite common in the languages of the world, but it is also a complex topic that would require a separate treatment.

2.1 *Third person*

The main conceptual sources for the genesis of new third person pronouns are listed in (7).

(7) *Sources of third person pronouns*[11]
   (a) Spatial deixis (demonstratives)
   (b) Nominal concepts
   (c) Intensifiers

We will now look at each of these sources in turn.

2.1.1 *Spatial deixis*

Perhaps the most common source for third person pronouns is provided by markers of spatial deixis, more precisely by demonstrative pronouns.[12] Third person pronouns are frequently created by means of a process where in specific contexts demonstrative pronouns are grammaticalized to third person markers (Heine & Reh 1984: 271; Diessel 1997, 1999b; Klausenburger 2000). Note that the conceptual source of grammaticalization is a demonstrative used as a pronoun, rather than as a nominal attribute (Diessel 1999b: 115). Relative spatial distance does not appear to be a decisive factor

[11] Drawing attention to the work of Johannes Helmbrecht (in particular Helmbrecht 2004), an anonymous *JL* referee points out that classifiers provide an additional source for third person pronouns e.g. in various Mixtecan and other Amerindian languages.

[12] An anonymous *JL* referee observes that it is 'odd to claim that 3rd person pronouns come from spatial deictics because in fact they usually come from forms marking discourse deixis (i.e. anaphoric demonstratives)'. While there are some examples that are in support of this hypothesis, we are not sure whether this is always the case. Evidence from sign languages, as well as from languages such as Korean (see below), suggests that it is in fact spatial deixis that provided the conceptual source for the rise of some third person pronouns (cf. Diessel 1997, 1999a, b). In any case, both are part of the following general pathway:

deictic demonstrative > anaphoric demonstrative > third person pronoun

in the process, even though there are more instances of the development that involve distal than proximal demonstratives.

At the initial stage of the process, therefore, the two are hard to separate in that the function of the relevant pronoun is ambiguous between that of a demonstrative and a personal pronoun in many contexts. The Uto-Aztecan language Cora appears to present such a stage in that 'all third person free pronouns are demonstratives. In the role of pronouns, then, demonstratives show up as subjects, direct objects, and objects of postpositions' (Casad 1984: 247). Another example is provided by the Australian Pama-Nyungan language Yindjibarndi, where Wordick (1982: 71) observes that all of the third person pronouns are also used as demonstratives, and Lewis (1985: 67–68) reports that in Turkish, the demonstrative pronoun *o* is used as a third person singular absolutive.

A similar situation appears to have existed in Ancient Egyptian, where the proximal demonstrative pronoun *pw* 'this' was also used as a general third person pronoun ('he', 'she', 'it', 'they'; Gardiner 1957: 85f., 103; Heine & Kuteva 2002). Finally, in earlier Korean, the demonstratives *i* 'this' (speaker-proximal), *geu* 'that' (hearer-proximal), and *jeo* 'that' (speaker/hearer-distal) were used as third person pronouns (Song 2002; see Section 2.2.2). At a more advanced stage, the demonstrative and the personal pronoun become separate functional categories, in that the latter undergoes desemanticization by losing its meaning of spatial deixis, decategorialization, and erosion.

That this is a development of crosslinguistic significance is also suggested by the fact that instances of it are also found in pidgins and creoles. In early Eastern Australian Pidgin English (EAPE) there are sporadic occurrences of *dat* (< English *that*) as a third person pronoun, e.g.:

(8) **Dat** make all black pellows get plentybark.
'He made the Aborigines collect a lot of bark.' (Baker 1995: 10)

And in the English-based creole Sranan of Surinam, the distal demonstrative *da* (historically derived from English *that* > *dati* > *da*) 'that' is used as a weak third person pronoun 'it' (Arends 1986).

2.1.2 *Nominal concepts*

The second major historical source of third person pronouns is provided by nouns. Heine & Song (2010) distinguish the kinds of nominal sources listed in (9).

(9) *Domains of nouns recruited for the grammaticalization of personal pronouns*
(a) Honorific expressions used for high-ranking personalities
(b) Terms for social status distinctions
(c) Kin terms

(d) Terms for professions
(e) Generic nouns for human beings, or based on distinctions of sex and age

Of these, it is most of all (9e), that is, nouns meaning 'man' or 'person' (Heine & Kuteva 2002: 112–113, 209–210) that tend to develop into personal pronouns. This pathway of grammaticalization has been documented most of all in Africa. The Central Sudanic language Lendu has suppletive forms for the noun 'man'. The singular stem *ke* 'man' changes to *ndrú* or *kpà* 'people' in the plural, and this suppletism has been retained in the grammaticalization from noun to personal pronoun – the third person singular pronoun is *ke*, and the third person plural pronoun *ndru* or *kpa*, all lacking tonal features (see (5b) above):

(10) *Lendu* (Central Sudanic, Nilo-Saharan; Tucker 1940: 392)
ma-zhi **ndru**. **ke** zhi **kpa**.
1SG-love 3PL 3SG love 3PL
'I love them.' 'He loves them.'

In the Central Khoisan language ǁAni of Botswana, the noun *khó*(*e*)*-mà* (person-M.SG) 'male person', 'man' has been grammaticalized to a third person masculine singular pronoun *khó*(*e*)*mà* 'he', which is undergoing erosion to *khóm*, e.g.:

(11) *ǁAni* (Heine 1999: 28)
… kánà **khóm̀** hin-ǁòè kx'éí-hɛ̀.
because he do-HAB manner-F.SG
'[The crocodile catches her] because this is the way he (=the crocodile) does it.'

Similarly, the Ubangi language Zande of Zaire has a noun stem **ko* 'man', 'male' which appears to have given rise to the masculine personal pronoun *kɔ* (Heine & Reh 1984: 223; Claudi 1985). The Nilotic Southern Lwoo languages of Uganda illustrate two different stages in the development from noun to personal pronoun. In two Ugandan languages, the noun for 'people' has developed into an anaphoric third person plural pronoun 'they', but whereas in Alur the form *jɔ̀* means both 'people' and 'they', in Adhola the pronoun *jó* 'they' is only diachronically a noun for 'people' (Heusing 2004: 218).

A further example is provided by Korean. We noted above that the demonstratives *i* 'this' (speaker-proximal), *geu* 'that' (hearer-proximal), and *jeo* 'that' (speaker/hearer-distal) were used as third person pronouns in earlier Korean; but since the end of the 19th century, when Korean was standardized as a written language, the demonstratives were reinforced by the noun strategy, whereby an appropriate head noun (e.g. *saram* 'person')

Level of honorification	Form	Approximate literal meaning
Highest	i seonsaengnim	'this person'
High	i bun	'this person'
Middle	i saram	'this person'
Lower middle	i chingu	'this fellow'
Low	i nom	'this fellow'
Low, abusive	i jasig	'this creature'

Table 1
Korean forms corresponding to third person pronouns in other languages (illustrated with the speaker-proximal demonstrative *i* 'this'; Song 2002: 10).

was added; Table 1 lists the various head nouns according to the level of honorification in Standard Korean.[13]

A wealth of third person pronouns derived from nouns expressing distinctions relating to social role relations can be found in Southeast and East Asia. In the Austro-Asiatic language Khmer, the form *kee:* serves not only as a noun for 'person' but also as a third person pronoun of neutral social status and as an indefinite pronoun (John Haiman, p.c.), and the Thai item *phráʔoŋ* is not only a person marker for second and third person, referring to high ranking persons, but also a noun or classifier for 'high ranking royalty'. Cooke (1968: 17) says that he is listing it 'as a personal pronoun because pronoun usage predominates over unambiguous noun usage'. And in Vietnamese, the item *ho* is a noun for 'family' but also a pronoun 'they' used when speaking of a group of persons who are neutral as to respect or deference, and *Nguời* is on the one hand a noun or classifier and on the other hand a third person marker used when speaking of gods, kings, or highly respected persons (Cooke 1968: 114).

2.1.3 *Intensifiers*

A third important source for third person pronouns is provided by what we refer to in a loose sense as 'intensifiers'. With this term we are referring to three kinds of pronominal concepts, namely reflexives, intensifiers, and identitives (or identity pronouns, 'the same').[14] What the three have in common is that they all presuppose some entity whose referential identity

[13] In addition, a distinction is made by means of verbal suffixes between hearer-honorific, speaker-honorific, speaker–hearer-honorific, and non-honorific reference (Song 2002: 11).

[14] The term 'intensifier' is also associated with a number of different notions in linguistics. We are using the term here in the sense defined by König & Siemund (2000), and described below.

has been established in previous discourse, and that they tend to undergo the same kind of grammaticalization process. In some of the existing literature, reflexive forms (e.g. *John killed* ***himself***) and intensifiers (*John* ***himself*** *killed her*) are not distinguished terminologically, even if intensifier pronouns tend to be referred to as 'emphatic reflexives'. It therefore remains frequently unclear which of the two is involved in the creation of personal pronouns; it would seem, however, that in most such cases it is intensifiers, rather than reflexives, that are used; more research is required on this point.

One example of an intensifier source is provided by the Basque identity pronoun *ber-* 'same, -self'.[15] Grammaticalization induced by contact with Romance languages appears to have affected the system of (independent) personal pronouns in Basque, being held responsible by Haase (1992: 135–137) for the fact that *ber-* is on the way to developing into a third person pronoun; we will return to this issue in Section 3.

Intensifiers in the Ethio-Semitic language Amharic are commonly formed by means of a noun phrase consisting of the noun *ras* (Ge'ez *rəs*) 'head' plus a possessive modifier, which appears to have provided the basis e.g. for a range of third person pronouns (Praetorius 1879: 119f.; Hartmann 1980: 273; Zelealem Leyew and Ulrike Claudi, p.c.). One may also mention Turkish, which appears to have exploited its intensifier in certain contexts to express third person reference:

(12) *Turkish* (Siewierska 2004: 226)
Kendi-si opera-ya git-ti.
self-3SG opera-DAT go-PAST
'He (respectful) has gone to the opera.'

That the development from intensifier to personal pronoun is in fact a process of grammaticalization is suggested by the fact that, first, it is unidirectional. Crosslinguistically there are a number of languages where intensifier ('-self') or identity forms ('the same') have given rise to personal pronouns, while there is so far no evidence for a development in the opposite direction. Second, this development can be described in terms of desemanticization (see Section 1.2 above), whereby the specific intensifier or identity semantics is bleached out – with the effect that third person reference, that is, a schematic deictic concept, is the only semantic function that is left.

2.1.4 *Other sources*

Demonstratives, nouns (or noun phrases), and intensifiers are not the only sources for third person pronouns. Another concept that can also be grammaticalized to a third person pronoun is 'the other(s)'. This has happened in

[15] The development from Ancient Greek *autós* '-self' to Modern Greek *aftós* is proposed by an anonymous *JL* referee as being another example.

some French-based creoles (see Section 2.2.1 below), as well as in Khmer, where the form *kee* 'other(s)' appears to have given rise on the one hand to a number-neutral third person pronoun and on the other hand to an expression for indefinite agents, much like French *on* (John Haiman, p.c.).

Once a third person pronoun has evolved, this is not necessarily the endpoint of its development. It may develop further into a second person pronoun, as we will see in the next section. But it may as well lose its pronoun status and end up as an agreement marker on the verb (Givón 1984: 353–360; see also Diessel 1999b: 120).

2.2 *Second person*

In this section we are confined most of all to expressions for singular referents. This however does not mean that plural reference is irrelevant, as we will see below. Second person plural pronouns are semantically complex, they may denote two or more addressees, or one addressee plus at least one nonparticipant (cf. Heath 2004: 1008), and in a number of cases, this complexity can be accounted for with reference to the respective diachronic sources of these pronouns.

Second person singular pronouns belong to the diachronically most stable parts of grammar. Nevertheless, Hagège (1993: 114f.) argues that presumably in the majority of the languages of the world there is one or more alternative form for them. Of all concepts of personal deixis, second person pronouns appear to have the largest range of diachronic sources. The sources listed in (13) appear to be crosslinguistically the most common ones:

(13) *Sources of second person pronouns*
  (a) Third person pronouns
  (b) Intensifiers
  (c) Nominal concepts
  (d) Spatial deixis
  (e) Plurification

A phenomenon that plays quite some role in the growth of many second person singular pronouns concerns PLURIFICATION.[16] Plurification concerns the use of plural pronouns, most commonly of second person plural pronouns, to refer to single persons;[17] it can be understood as a strategy whereby the use of a plural pronoun is extended to also denote singular referents with the intent to express distinctions in social status, politeness, and/or

[16] In Heine & Song (2010), the term pluralization was used instead. We are grateful to an anonymous *JL* referee for pointing out that 'pluralization' is misleading and for proposing 'plurification' as one possible alternative, and we have adopted this term.

[17] Navajo has been reported to use a dual rather than a plural form to express degrees of respect or social distance (Head 1978: 158).

social distance; following a widespread usage, we will refer to such distinctions summarily with the term honorification.[18] A paradigm example is provided by French, where the plural pronoun *vous* 'you.PL' was extended to singular referents; for a detailed discussion of plurification, see Heine & Song (2010).

2.2.1 *Third person pronouns*

Third person pronouns are crosslinguistically perhaps one of the most common sources for second person pronouns (see e.g. Head 1978: 167–171). It is especially third person plural pronouns, rather than singular pronouns, that are grammaticalized to second person pronouns. Plural forms have played quite some role in e.g. the history of deictic marking in German (see Simon 1997), the Modern High German polite address pronoun *Sie* 'you' (<'they') being a modern reflex of this grammaticalization source; see Section 2.2.2 for more details. In fact, German provides a paradigm example of the third-to-second strategy. The way this affected the structure of personal deixis in the German of the 17th century can be shown with the following quotation:

> The seventeenth century was the heydey of honorific flourishes and the position of the 3rd person pronoun was strengthened by the continual employment of such expressions as *Euer Gnade* 'Your Grace'. But more than this, plural phrases like *Euer Gnaden* 'Your Graces' came into fashion as hyper-polite modes of address. This development provoked a corresponding employment of the 3rd pl. *Sie* as the most respectful pronoun of all; one could now address a person as 'They'. By the second half of the eighteenth century the plural *Sie* had established itself. (Lockwood 1968: 62)

Some consequences of this process are discussed below. The process may be due either to a straightforward shift from third to second person, or to the fact that the third person pronoun at first refers anaphorically to a nominal address form. It would seem that the latter is more common:[19] Third person pronouns developing into second person pronouns tend to start out as anaphoric pronouns referring to nominal expressions of address but serving to address second person referents, in a way that can be illustrated with the following example from German. In some upper-middle–class restaurants the customer may not be surprised to be addressed by the waiter as in (14),

[18] A perhaps more appropriate, but also more clumsy term for distinctions in social meaning would be 'politeness, respect and/or social distance' (see Head 1978: 153).

[19] Conceivably, the anonymous *JL* referee is correct who suggests that an anaphoric route applies perhaps not only to the latter but to all shifts. More data are required on this issue.

where the third person pronoun *er* 'he' denotes second person reference but refers anaphorically back to the address form *der Herr* 'the gentleman'.[20]

(14) *German*
Was wünscht der Herr? Hat er sich schon entschieden?
what wants the gentleman has he himself already decided
'What would you like, Sir? Have you decided already?'

Whether the noun phrase is a singular or a plural form is likely to determine whether the resulting second person pronoun will have singular or plural reference. Examples from other European languages, such as Italian, Danish, Norwegian, or Swedish, suggest in fact that third person pronouns developing into second person pronouns more commonly originate as third person anaphoric markers for honorific noun phrases used as address forms (Head 1978: 168; Helmbrecht 2004, 2005a).

This pathway, though, is certainly not the only one to be held responsible for the third-to-second strategy. A number of French-based creoles have experienced a development where a form etymologically derived from French *les autres* 'the others' has given rise to third person plural pronouns (> *zot*), see (15), and further to second person plural pronouns, see (16).

(15) *Mauritian* (French-based creole; Syea 1996: 176)
liv la ki zot ti lir.
book DEF that they TNS read
'It was the book that they read.'

(16) *Guyanese Creole French* (Corne 1971: 92)
mo pa uer zot pies.
(I NEG see you.PL piece)[21]
'I didn't see anyone of you (plural).'

2.2.2 *Intensifiers*

Intensifiers, reflexives, and identity pronouns (or identitives) are cross-linguistically a common source for second person pronouns, especially in languages spoken in India and generally in Southeast Asia (see Head 1978: 179ff.). An example of an intensifier is found in (17) and of an identifier pronoun in (18), involving, respectively, the Korean intensifier/reflexive[22]

[20] This usage appears to be gradually disappearing; it can be observed mostly with older waiters, and younger Germans may no longer be aware of it. For a similar example from modern French, see Head (1978: 168).

[21] Parentheses indicate that in this example, and some examples below, there are no glosses in the original and that the glosses are ours.

[22] The pronoun *jane* serves both as an intensifier and a reflexive.

jane 'oneself' and the 18th-century German identity pronoun *dieselben* 'the same ones'.

(17) *Southeastern Korean* (Song 2002)
jane eonje ga-lgeon-ga?
INT/2.ABS when go-FUT-INTR
'When will you go?'

(18) *18th-century German* (Simon 2003)
Ich bitte dieselben.
I ask the.same.ones
'I ask you.' (Maximally polite form of address)

The crucial role played by intensifiers in the development of personal deixis marking can be illustrated with the following example from Korean.[23] In accordance with Song (2002) we distinguish between three levels of honorification: a high level, a middle level, and a low level.[24] At least since the 16th century, the form *jeo* was used as an intensifier/reflexive form of all honorific levels. Towards the end of the 16th century, three new intensifier/reflexive forms arose: *dangsin*, which became the marker of the high level, and *jagi* and *jane*, which became the middle honorific level forms, with *jeo* being degraded to a low level form. Around the end of the 19th century, *jeo* developed into a first person pronoun of the low honorific level.

The form *dangsin*, introduced into the language as a high level intensifer/reflexive towards the end of the 16th century, acquired an additional use as a second person pronoun of the high level around the end of the 18th century. This polysemy of *dangsin* continued up until around the mid-20th century, when *dangsin* was discontinued as an intensifier/reflexive and devaluated[25] to a second person pronoun of the middle level.

The use of *jane* was extended from a middle level reflexive to second person pronoun of the middle level in the mid-17th century, and shortly thereafter its intensifier/reflexive use was discontinued. As a second person

---

[23] Intensifiers/reflexives have also experienced other kinds of grammaticalizations, such as developing into impersonal pronouns. Such developments are ignored here.

[24] Discussion is confined to singular uses of personal pronouns. Note further that, while Song (2002) is concerned with the development of the full paradigms of pronouns, we are confined here to the development of individual items from reflexives to personal pronouns. We are also ignoring the fact that intensifier/reflexive pronouns themselves can be traced back to earlier noun phrases (see Song 2002 for details), so that we are dealing with a more general grammaticalization of the kind:

noun phrase > intensifier/reflexive > personal pronoun

According to an anonymous *JL* referee, it is intensifiers rather than reflexives that are involved here. While this is in fact possible, our database is not conclusive on this point.

[25] Concerning the term 'devaluation', see Song (2002). Devaluation typically means that an address form loses some or all of the honorific significance it once had.

pronoun, it was devaluated to a low level pronoun around the mid-20th century, and in modern Standard Korean it is restricted to addressing people younger than the speaker, although it appears to be falling into disuse (Song 2010). The second middle level intensifier/reflexive *jagi* experienced a similar development as *jane*: It also acquired the function of a second person pronoun of the middle level in the second half of the 20th century and has remained polysemous between its reflexive and deictic uses up to the present, even though its function as a reflexive was generalized to all three levels of honorification.

To sum up, within the last five centuries there were four developments from intensifier/reflexive to deictic marker in Korean, one giving rise to a first person pronoun (*jeo*), while the remaining reflexives developed into second person pronouns.

Korean is by no means an isolated case. In Khmer there is no genuine second person pronoun, but the intensifier and reflexive form *aeng* 'oneself', typically in combination with *kluan* 'body', has acquired the meaning of a pronoun for a second person of neutral status ('you (all)').[26] The Thai second person marker *tuá*, literally meaning 'body, self', is (i) an affectionate or intimate term, used for speaking to equal or to inferior not older than the speaker, and (ii) used by a child or young woman speaking to an intimate (sibling, friend, fiancé, husband) in anger, expressing a sense of injury, and implying a disavowal or impersonalization of the close relationship (Cooke 1968: 17).

Another language where intensifiers or reflexives appear to have provided a productive pattern for personal pronouns is the Semitic language Amharic of Ethiopia. Intensifiers in Amharic are commonly formed by means of a noun phrase consisting of the noun *ras* (or Ge'ez *rəs*) 'head' plus a possessive modifier. It would seem that this construction provided the basis for e.g. the second person singular respect pronoun *ərs^{w}o* 'you' (<**əras-wo* 'your (respectful) head'), but also for a range of third person pronouns, as we saw in Section 2.1.3 above.

Furthermore, Siewierska (2004: 227) mentions Tetelcingo Nahuatl, where there are even two reflexive markers that can be used jointly: On the one hand the prefix *mo-*, derived from the third person reflexive, and the reflexive marker *-cinow*, used in the case of second person honorific reflexives:

(19) *Tetelcingo Nahuatl* (Siewierska 2004: 227)
To-**mo**-kokoh-**cinow**-a.
2.SG-REFL-hurt-REFL.H-PRES
'You (honorific) are sick.'

[26] Typically in conjunction with *kluan* 'body'. That the reflexive and intensifier meanings constitute the source of grammaticalization is suggested by the fact that this is also the meaning of the etymologically related form in Thai (John Haiman, p.c.).

Compared to the Korean, Amharic, or Khmer situations, the development of the German identity pronoun *dieselben* 'the same ones' to a second person pronoun in the 18th century is suggestive of a process that was restricted in a number of ways. The pronoun surfaced in the middle of the 18th century as an anaphoric pronoun of the highest level. Thus, around that time the influential grammarian Johann Christoph Gottsched describes the following four levels of politeness that were distinguished: *dieselben* = *überhöflich* ('super polite'), *Sie* = *neuhöflich* ('new polite'), *er* = *mittelhöflich* ('middle polite'), *ir* = *althöflich* ('old polite'), and *du* = *natürlich* ('natural') (Gottsched [1762]1970: 280). But *dieselben* had a relatively short lifespan, declining in the 19th century, even if it was still mentioned occasionally around 1900. Furthermore, it was largely confined to plural addressees and to uses as an anaphoric marker referring to some antecedent title of dignity (Simon 1997: 274–275), and it was never conventionalized to a full-fledged second person pronoun (Simon 1997: 268ff.; 2003: 127).

When grammaticalized to a second person pronoun, an intensifier is likely to develop into an invariable form that may no longer refer to different categories of personal deixis. The 'reflexive' form *ap* of Hindi and Urdu, for example, occurs in all categories of person and number; as an address form, however, where it expresses the most respectful pronominal form of address, it has only second person reference (Head 1978: 180).

Intensifiers may combine with plurification (see Heine & Song 2010) to form pronouns of respect or social distance. In Bengali, the second person singular pronoun is an 'inferior' address form while the second person plural pronoun is the 'common' address form. But the most respectful pronoun is *apnara*, which is a grammaticalized plural form of the 'reflexive' *apni* (Head 1978: 180).

Siewierska (2004: 226) found that outside the Indian subcontinent the use of reflexives as polite or respectful forms of address is not particularly common. Nevertheless, we have seen that there are a number of other societies outside that subcontinent that have also drawn on intensifiers for functions of personal deixis.

The exact pathways leading from 'intensifier' categories to personal pronouns are far from clear; a more detailed analysis of these categories and of the semantic changes leading to grammaticalization is urgently needed.

2.2.3 *Nominal concepts*

Shibatani (1990: 371–372) observes that etymologically, most of the personal pronouns in Japanese 'derive from regular nouns, e.g. *watakusi* "I" (<"private (thing)"), *kimi* "you" (<"emperor"), *anata* "you" (<"yonder")'. When grammaticalized to second person pronouns, nouns are likely to occur in possessive constructions, consisting of a noun and a possessive modifier, used as honorific address forms which assume the function of

second person pronouns. Mechanisms of grammaticalization involved in the process concern especially the loss of nominal meaning (desemanticization) and of phonetic substance of the forms concerned (erosion; see Section 1.2 above).

Desemanticization has the effect that the meaning of the nominal possessive construction is bleached out, resulting in the rise of a marker of second person deixis used for respect and/or social distance, decategorialization means that the ability to take nominal inflections, derivations and modifiers disappears, while erosion leads to a loss in phonetic substance of the expression concerned. Paradigm examples can be found in some Romance languages: Spanish *Vuestra Merced* and Portuguese *Vossa Mercê*, both meaning 'Your Grace', were grammaticalized to second person polite pronouns (*Usted* and *Você*, respectively), or Dutch *Uwe Edelheid* 'Your Nobility', which also developed into the second-person polite form *U* (see Head 1978: 185ff.; Siewierska 2004: 224). In accordance with their origin as noun phrases, the resulting forms are likely to exhibit third person agreement, even though they are now unambiguously markers of second person deixis (see Comrie 1975). Whether, or to what extent, the head nouns functioning as possessees in such constructions can be regarded as semantically parasitic heads (Levine 2010) cannot be determined on account of lack of appropriate data but should be an issue of future research.

Portuguese *Vossa Mercê* began to be used as a polite form of address for the king around 1460, but was soon devaluated. Around 1490 its use was extended as an address form for dukes, in the 16th century it was further extended to the bourgeois population, and in the 18th century it replaced the earlier *vós* as a polite and ceremonial form of address for singular referents.

A similar process took place in Romanian, where the noun phrase *Domnia Ta* 'Your Grace' was grammaticalized to the second person address form *dumneata* (also: *Domnia Voastră* 'Your (plural) Grace' > *dumneavoastră*). This process was presumably associated with the rise of the Romanian feudal society between the 10th and the 14th centuries, and the synthetic form of the pronoun is documented already in the 16th century (Merlan 2006: 222–226). We noted in Section 1.2 that erosion is a frequent but not a necessary parameter of grammaticalization, and the present examples of the evolution of nominal honorific expressions illustrate this fact: While the developments in Spanish, Portuguese, or Dutch show massive erosion, there was hardly any erosion in the grammaticalization of the Romanian pronouns.

In a comparable fashion, speakers of some languages of Southeast Asia[27] have extended the use of nominal expressions designed by the speaker to

[27] As we note in section 2.3.1.2, languages where this is said to have happened include Burmese, Javanese, Khmer, Malay, Sundanese, Thai, and Vietnamese (Cooke 1968; Head 1978: 187). In many cases, however, this usage has not led to the rise of fully grammaticalized forms of personal deixis.

humble himself or herself in self-reference to markers of first person reference, and nominal expressions signaling respect to markers of second person address. For example, the second person honorific pronoun *tuan* 'you' of Indonesian is said to originate in an Arabic loan meaning 'master', and Old Japanese *kimi* 'emperor' or 'lord' developed into a second person singular pronoun 'thou' (Shibatani 1990: 371–372). Note further that the Khmer noun *neak* 'person' is conscripted for use as a second person pronoun 'you (younger than speaker)' (John Haiman, p.c.), and the Thai item *phráʔoŋ* is not only a person marker for second and third person, referring to high ranking persons, but also a noun or classifier for 'high ranking royalty' Cooke (1968: 17; see Section 2.1.2).

Table 2 (on next page) lists examples from Southeast and East Asia that lend support to what we propose to call the ASYMMETRIC-ROLE MODEL OF SPEAKER–HEARER INTERACTION. According to this model, the speaker understates his or her own social status, describing himself or herself as being of low status while treating the hearer as a personality of high status, deserving a respectful form of address.[28]

In quite a number of languages there is a secondary strategy of using kin terms or terms for professions (see (9c, d) above) as weakly grammaticalized forms of second person pronouns. In Korean, for example, *elusin* is employed rarely as a deferential second person pronoun, referring to a respected male of over sixty years of age and as a result, reference terms such as *apeci* 'father', *sensayng-nim* 'teacher', and *sacang-nim* 'company president' are used as pronominal substitutes (Sohn 1999: 208).

2.2.4 *Spatial deixis*

Being a language where avoidance of direct address plays an important role, Korean has developed a range of different expressions for personal deixis. Among them are the following expressions for distal space which can serve as third person pronouns: the noun phrases *jeo jjog/pyon* 'that side', *jeo gos* 'that place', and the locative adverb *jeogi* 'there' (Song 2002; see Section 2.3.1.3). In addition, the distal adverb *jeogi* 'there' may be used in the function of a second person pronoun. But such expressions are for the most part instances of incipient grammaticalization, not having been incorporated into the paradigm of personal pronouns (suggestive of a bridging context; see Section 3, Table 6 below), as is also the case in a number of other languages where distinctions of spatial deixis provide a contextually available alternative, that is, they are not conventionalized to regular forms of personal deixis.

In some cases, however, spatial deixis has given rise to fully grammaticalized second person pronouns. The Japanese second person pronoun *anata*,

[28] It would seem that manifestations of this model can also be found in societies beyond the ones mentioned here; for example, in many linguistic communities it is perfectly normal for the speaker to say 'I am stupid' but far less socially appropriate to say 'You are stupid'.

Language	First person singular	Nominal source	Second person singular	Nominal source
Burmese	tyunvdov or tyanov	'royal slave'	minx	'king'
Indonesian	saya	sahaya 'servant' (<Sanskrit)	tuan	'master' (<Arabic)
Japanese	boku	'slave'	kimi	'emperor, lord'
Khmer	knyom	'servant'	loːk ('2.higher status')	'monk'
Thai	khâá	'servant, slave'	naaj́	'master, mistress'
	khâaphacâw	'lord's servant'	phráʔoŋ ('2SG, 3SG, used for high-ranking persons')	'high-ranking royalty' (noun or classifier)
	khâá-phráphúd́thacâaẃ	'Lord Buddha's servant'		
Vietnamese	tôi, tə́	'servant'		

Table 2
Some nominal sources for first and second person markers. (See the text for information on constraints and the social environment in which these forms are used.)

for example, originates in a spatial deictic noun signaling roughly 'over there' (or 'that part') in Late Old Japanese, having been used as a third person marker ('person over there') in Early Modern Japanese, before it shifted to a second person pronoun around 1750 (Traugott & Dasher 2002: 230). This example is suggestive of a chaining process whereby expressions for spatial deixis give rise first to third person markers which again develop into second person pronouns.

2.2.5 *Plurification*

Finally, there is another important strategy that provides a source for new categories of second person deixis, namely plurification. With this term we refer to a process whereby second or third person plural pronouns are also used for (honorific or polite) second person singular address. What distinguishes this process from other processes discussed in this paper is that it is restricted to one and the same paradigm, in that one category of person marking turns into another category of person marking; we will return to this issue in Section 3. A number of languages in Europe have developed new forms of second person singular address; in English the process has proceeded further, with the effect that the new pronoun *you*, formerly a honorific pronoun, has replaced the old pronoun *thou* (Brown & Gilman 1968; Helmbrecht 2004, 2005a).

Plurification is discussed in detail in Heine & Song (2010; see also Siewierska 2004: 216–221 and Sections 3 and 4 below); one example may therefore suffice to illustrate the process. The process looked at occurred within the last fifty years in a modern African society. Sango, a Ubangian language of the Niger-Congo family, is the national language of the Central African Republic, being an official language side by side with French. One of the changes that this language underwent recently concerns the rise of a new functional category via the extension of the second person plural pronoun *álà*, resulting in a new deferential second person singular pronoun (Samarin 2002). The use of the new category, which must have arisen in the 1960s is flexible: Except for children under 13 years of age, essentially any Sango speaker can give *álà* to anyone else instead of the traditional second person singular pronoun *mò*, but female addressees are generally more likely to receive *álà* than males, and the probability that grandparents and parents receive *álà* is the highest. Being a young category, it has not yet been fully conventionalized; its use remains largely optional, being determined to some extent by social and demographic variables.

2.3 *First person*

Since singular and plural forms of first person pronouns mostly follow different pathways of grammaticalization (see (1d) above) we are treating the two separately in the present section.

2.3.1 *First person singular pronouns*

There is only limited information on the origin of first person singular pronouns (see e.g. Blake 1934, Cooke 1968), and where there is information, it concerns oppositions in deictic space and social status.

2.3.1.1 *Spatial deixis*

Claude Hagège characterizes the spatial deixis strategy serving the expression of first and second person pronouns in the following way:

> [There are] languages which use spatial adverbs with the meaning of personal pronouns: Japanese *kotira* 'here' often refers to the speaker, Vietnamese *Hây* 'here' and *Hây* (or *Hó* 'there') are used with the meanings 'I' and 'you' respectively when one wants to avoid the hierarchical or affective connotations linked to the use of personal pronouns. (Hagège 1993: 216–217)

As we observed in Section 2.2.4, Korean also has recruited distinctions of space for the expression of personal deixis, using proximal demonstrative attributes with locative nouns. Thus, *i jjog/pyon* 'this side' or *i gos* 'this place', as well as the locative adverb *yeogi* 'here', tend to be used as first person pronouns; consider example (20), which illustrates the significance of spatial deixis in encoding the contrast between first and second person referents.[29]

(20) *Standard Korean* (Song 2002: 14)
i jjog-eun gwaenchan-eunde geu jjog-eun eotteo-seyo?
this side-NOM good-CONN that side-NOM how-END
'I am OK, and how about you?'

In Thai, the demonstrative form *nîi* 'this (one)' is used as a first person singular pronoun by certain young women to male non-intimate equals (Cooke 1968: 14); consider also, for comparison, the use of spatial deixis in current colloquial English:

(21) *Colloquial English*
A: I hate this place!
B: Same **here**! (='So do I.')

[29] An anonymous *JL* referee, citing Margetts & Austin (2007), points out that in quite a few languages, words for 'hither' can be used for first person reference, even though they may not be incorporated into the pronominal paradigm – in other words, they appear to represent an incipient stage of grammaticalization (see Section 3 below).

2.3.1.2 *Nominal concepts*

Nouns used as a source for first person singular pronouns are found especially in some languages of Southeast Asia, such languages being Burmese, Javanese, Khmer, Malay, Sundanese, Thai, and Vietnamese (Cooke 1968; Head 1978: 187). Underlying this usage there appears to be the asymmetric-role model of speaker–hearer interaction that we proposed in Section 2.2.3 above, whereby nominal expressions are designed by the speaker to humble himself in self-reference, turning into markers of first person reference, and nominal expressions signaling respect into markers of second person address. The examples that we listed in Table 2 were but a selection of the expressions used in languages such as Burmese and Thai to encode first and second person address in terms of concepts that are in accordance with this model (see Cooke 1968 for more examples).

Already at the stage of Classical Malay, the literate noun *sahāya* (which is said to ultimately come from Sanskrit *sahāya* 'assistant') was polysemous between 'servant' and 'I (humble/polite)'. Later on, the nominal meaning was lost and *sahāya* underwent erosion to become *saya*, the first person pronoun for neutral, non-familiar 'I' in modern Indonesian (Lehmann 1995, Cysouw 2003: 13).

Among the many nominal forms used for first person address in Thai there is the noun *khâá* 'servant, slave', employed as a first person singular pronoun when speaking to inferiors, while *khâaphacâw* 'lord's servant' is used in public address for first person reference (Cooke 1968: 13). In Vietnamese, the noun *tôi* 'servant' (taking the plural marker *các* 'group') is also used as a person marker for 'I, me', and in this case the plural marker is *chúng* 'they'. The person marker is, however, not ordinarily employed between adults and children or vice versa, or between close blood kin; but otherwise its use is general and fairly neutral. Furthermore, the noun *tớ* 'servant' was formerly used especially in North Vietnam as a first person pronoun speaking to one's schoolmates (Cooke 1968: 113). And finally, we noted above that in Japanese the noun *boku* 'slave', a Chinese loanword, gave rise to the first person pronoun *boku*.

2.3.1.3 *Intensifiers*

The two sources discussed above account for many but not all pathways of grammaticalization leading to the rise of first person singular markers. Once again, we find cases of intensifiers or reflexives among the sources, like Korean *jeo*. Around the end of the 19th century, the Korean intensifier/reflexive form *jeo* was extended to use in the domain of personal deixis, becoming a first person pronoun of the low honorific level, and it has retained this function up to the present, with the traditional first person pronoun *na* being reduced to use at the middle honorific level (see Section 2.2.2 above).

In Standard Japanese, the reflexive marker *jibun* (or *zibun*) may be used for 'I' as a first person pronoun, but grammaticalization has gone further in the Kansai dialect area of Japanese, where the reflexive marker *jibun* is used much more frequently in this capacity, generally acting as one of the ordinary personal pronouns for 'I' (Hinds 1986: 124); the following example is taken from the usage in Osaka:

(22) *Japanese* (Hinds 1986: 263)
isshookenmei **jibun** ga mooshikomi kaite mootekiteta yaro.
diligently I NOM application write took COP
'I wrote out the application diligently and took [it there].'

The grammaticalization of reflexive *jibun* to a first person marker was, however, not approved by everybody in Japan. After the end of the American occupation, the government document *Kore kara no keigo* ('Honorific Language from Now On') was officially released. This memorandum of 1952, which can be seen as a signal of Japan's transition to a modern nation state with a new orientation in social relations, proposed a number of new regulations on person markers, and one of them was geared at eliminating *jibun* as a person marker: 'The use of *jibun* "self" instead of [the first person pronoun] *watashi* should be avoided' (Coulmas 1992: 308).

There are also a few other languages, in Southeast Asia, where the use of intensifiers or reflexives includes that of first person reference, like Vietnamese *mình* 'body, self', which also serves as a first person term chiefly among females speaking to close intimate equals of either sex (Cooke 1968: 112). One may further mention the Thai first person pronoun *ʔàadtamaá* that is used by a Buddhist priest speaking to a layman or lower ranking or formally equal priest, which appears to be derived from a term meaning 'the self, the individual' (Cooke 1968: 11).

2.3.2 *First person plural pronouns*

As we observed in (1d) above, more than other deictic categories, first person plural pronouns are conceptually complex and tend to have suppletive forms (Forchheimer 1953: 65; Heath 2004: 1008). But not all languages have a distinct first person plural pronoun, and some languages, such as the Alacalufan language Qawasqar of Chile use the same pronoun for first person singular and first person plural (see Cysouw 2003). On account of the complex interaction of personal deixis and number of first person plural pronouns, analyzing their structure and grammaticalization behavior would require a study of their own (but see Cysouw 2003, 2005a, b).

2.3.2.1 *Nominal concepts*

There is one salient pathway leading to the genesis of first person plural pronouns, namely via nouns meaning 'person' or 'people' serving as a

Language	Noun	First person plural pronoun
Saafi-Saafi	ɓoʔ 'person'	ɓoo 'we'
Laala*	ɓoyyi 'person'	ɓi 'we EXCL'
Palor of Khodoba	ɓoo 'people'	ɓoo 'we'

*In Laala, the noun ɓ*oyyi* 'person' has also developed into an impersonal pronoun.

Table 3
First person plural pronouns in Cangin languages of Senegal and their lexical sources (Ursula Drolc, p.c.).

conceptual source. Most of the examples found so far stem from African languages. In the Central Sudanic language Ngiti of the Nilo-Saharan family, the noun *alɛɛ* 'person, people' appears to have given rise to a first person plural inclusive pronoun *àlɛ̀* 'we'. In the process of grammaticalization, the erstwhile noun was, in accordance with (5c) above, adapted to the phonological paradigm of pronouns, taking the same tonal pattern as the first and second person plural pronouns, and it underwent erosion, being shortened to *lɛ̀* or *l-* in fast speech (Kutsch Lojenga 1994: 195).

Another example is provided by the languages of the Cangin group of Senegal, belonging to the Atlantic branch of the Niger-Congo family. Three of the five Cangin languages appear to have experienced a similar grammaticalization whereby a noun for 'person' or 'people' has given rise to a first person plural pronoun; Table 3 presents the pronominal forms and their hypothesized lexical sources. This transfer cannot be traced back to Proto-Cangin, that is, it appears to have occurred independently in the different languages after the split of the proto-language.

African languages offer a wealth of further examples of the development from noun to first person plural pronoun. In the northern and western dialects of !Xun of the North Khoisan family, spoken in South Angola and North Namibia, the noun *djù* 'person' has given rise to a first person plural exclusive pronoun, see (23).

(23) *!Xun, W2 dialect* (North Khoisan, Khoisan; own field data)
djù- tcā má g‖à- ‖xˈāē.
I.PL.EX DU TOP stay together
'We two stay together.'

In the Mande language Kono of Sierra Leone, the noun *mɔ̀ɔ̀* 'man, person, people' (24a) has developed into a first person plural pronoun *mɔ̂* 'we (inclusive)', see (24b), where on the way from nominal meaning to marker of personal deixis the item concerned underwent erosion.

(24) *Kono* (Mande, Niger-Congo; Donald A. Lessau, p.c.)
(a) **mɔ̀ɔ̀** kúndú- nù
person short PL
'short people'
(b) **mɔ̂** dè án né.
we.IN mother EMPH here
'This is our mother.' (Historically: *'This is people's mother.')

Susu, another Mande language of Guinea, has developed its noun *mikhi* 'man', exemplified in (25a), into a first person exclusive plural pronoun *mukhu* 'we, us, our' (Friedländer 1974: 25; note that there is free variation in Central Mande between the high vowels *i* and *u*), illustrated in (25b).

(25) *Susu* (Mande, Niger-Congo; Friedländer 1974: 28)
(a) **mikhi** mundue?
(man which)
'which people?'
(b) **mukhu** khunyi
(we head)
'our heads'

In the Labwor language of northeastern Uganda there is an inclusive first person plural pronoun *jɔ̀* 'we (including you)'. As the evidence from closely related Western Nilotic languages suggests, this is a grammaticalized form of the noun *jɔ̀* 'people'. The Labwor pronoun is very weakly grammaticalized, meaning that it is never used as a subject pronoun and only occasionally used as an object pronoun, as in (26a), and only with a limited number of head nouns as a possessive attribute, as in (26b). Furthermore, there is considerable variation among speakers as to whether its use is allowed or disallowed.

(26) *Labwor* (Western Nilotic, Nilo-Saharan; own field data)
(a) én ɔ́nɛ̀nɔ̀ **jɔ̀**.
3SG see.PFV.3SG people
'He has seen us all.'
(b) ɛ̀thínɔ̀ **jɔ̀**.
children people
'our children (including yours)'

To conclude, there are examples from a range of African languages, covering three of the four African language phyla: The Niger-Congo family is represented with languages such as Kono, Susu, and the Cangin languages, Nilo-Saharan with Ngiti, Labwor and other Western Nilotic languages, and Khoisan with !Xun.

But there are also examples from other parts of the world, including European languages. A paradigm example is provided by Portuguese, where the feminine singular noun phrase *a gente* 'the people' has been

grammaticalized to a first person plural pronoun, to some extent also to a first person singular pronoun – a process that is said to have started in the 17th century.[30] In its lexical form, *gente* 'people' is a full-fledged noun, which can be inflected and modified; as a pronoun it has undergone decategorialization in its definite form (with the preposed article *a* 'the'), as is illustrated in (27).

(27) *European Portuguese* (Merlan 2006: 230)
A gente fala mais tarde.
(the people talk more late)
'We'll talk later.'

While *a gente* typically shows third person singular agreement, it has been adapted conceptually and morphologically to its new function to some extent, in that there are also quite a number of examples where there is first person plural agreement. The extent to which *a gente* has been grammaticalized can be seen in the fact that, at least in European Portuguese, it serves not only as a subject pronoun but also as a direct and an indirect object pronoun, a prepositional object pronoun, and even as a possessive attribute:[31]

(28) *European Portuguese* (Merlan 2006: 236)
Pode ficar na casa **da gente.**
can remain in.the house of.the people
'You can remain in our house (or with us).'

This new deictic is found not only in European Portuguese but rather occurs in all four continents where Portuguese is spoken. In all the varieties the noun phrase *a gente* 'the people' has been grammaticalized to a first person plural pronoun (Merlan 2006). As we will see in Section 3, in Brazilian Portuguese, *a gente* 'we' has – more than in European Portuguese – replaced the earlier pronoun *nós*, also serving both as a first person plural pronoun ('we') and as an impersonal pronoun ('one'; Albán & Freitas 1991; Schwegler 1993: 152; Merlan 2006: 235; Travis & Silveira 2009; Martelotta & Cezario in press).

And there is also evidence from a creole language. In the Spanish-based creole Palenquero of Colombia, the noun (*ma*) *hende* 'people', ultimately derived from Spanish *gente* 'people', has given rise to a pronoun (*h*)*ende*,

[30] In addition, *a gente* also shows uses of an indefinite pronoun corresponding to French *on* or German *man* (Merlan 2006: 233).

[31] An anonymous *JL* referee notes that one would expect that with increasing grammaticalization, the possibilities for syntactic position of *a gente* would decrease rather than increase. As a matter of fact, however, extension to new contexts (including new syntactic positions) is a central mechanism of grammaticalization (see (2) in Section 1.2 above and Table 5 in Section 3 below).

which serves both as an impersonal pronoun 'one' and as a first person pronoun 'we', see (29).[32]

(29) *Palenquero* (Spanish-based creole; Schwegler 1993: 152–153)
Hende a ablá pa bo rihá ri e ma kusa!
people/we TAM say for you leave alone this PL thing
'We have [already] told you to leave these things [=a machete] alone!'

Another example concerns the French impersonal pronoun *on* 'one', which historically goes back to Latin *homo* 'person, man'). In informal spoken French it has been grammaticalized to a first person plural pronoun *on*, which is said to be used by all French speakers, side by side with the inherited pronoun *nous* 'we'. As example (30) shows, it has not yet undergone extension to the extent that it is used in all contexts.

(30) *Colloquial Modern French* (Coveney 2000: 459)
C'est nous qu' **on** est les vainqueur-s.
(that.is we who we is the.PL winner-PL)
'It is us who's the winners.'

A final example concerns another Romance language, namely Romanian. In this language, at the eastern end of the Romance world, there is a pronoun *lumea*, which is historically derived from the feminine noun *lume* 'people, the people, human beings' plus definite article *-a*;[33] (31a) exhibits the lexical and (31b) the deictic use of this item:

(31) *Romanian* (Merlan 2006: 227)
(a) **Lumea** a râs.
(people.DEF have laughed)
'The people have laughed.'
(b) Hai, Mihaela, că **lumea** pleacă.
(come Michaela here people.DEF go.away)
'Come on, Michaela, we go away.'

As a personal pronoun for 'we' or 'you (plural)', *lumea* obligatorily has the definite article on it. Thus, like the corresponding form *a gente* in Portuguese it is decategorialized, no longer being inflected, and taking no modifiers other than the quantifier *toată* 'all (feminine singular)' and the locative adverb (*de*) *aici* '(from) here'. In grammars of written Romanian, *lumea* tends to be ignored. It serves as an inclusive first person pronoun when

[32] The meaning of (*ma*) (*h*)*ende* is frequently vague, in that it may be interpreted variously as a noun, an impersonal pronoun, or as a first person plural pronoun.

[33] Ultimately, Romanian *lumea* goes back to Latin *lūmen* 'light', which acquired the additional meaning 'world' under the influence of Slavic *světŭ* 'light, world' (Merlan 2006: 226).

the speaker and other persons form a *Solidar- oder Konsensgruppe* (Merlan 2006: 228, 231). Merlan points out that (*toată*) *lumea* as a deictic pronoun is a 'sporadic phenomenon', being restriced almost totally to use as a subject pronoun.

A common pathway underlying this grammaticalization process, even if presumably not the only one, appears to be the following: noun ('person' or 'people') > indefinite pronoun ('people in general') > first person plural pronoun. For example, in Brazilian Portuguese, *a gente* 'the people' was first used as an indefinite pronoun before it turned into a personal pronoun 'we' in the 19th century (Travis & Silveira 2009: 249).

2.3.2.2 *Discussion*

What such examples suggest is that there is robust evidence for a conceptual development from 'person' or 'people' to first person plural pronoun, whereby a deictic pronominal category is conceptualized in terms of a non-deictic entity, a concrete noun. As we saw above, the result may be a new pronoun for either inclusive reference ('we, including you') or exclusive reference ('we, excluding you'), or both; more research is needed on this issue. Initially, the process tends to be confined to contexts where the noun serves as the subject of the clause, before it spreads to contexts where it is also used as the object or as a possessive modifier; but once again there is at least one exception: In the Nilotic language Labwor, the noun *jɔ* 'people' has given rise to an inclusive object or possessive pronoun but not to a subject pronoun (see Section 2.3.2.1 above).

All the examples presented above are in accordance with the four parameters of grammaticalization introduced in Section 1.2. Extension, i.e. use in new contexts that suggest a deictic meaning, desemanticization, whereby the nominal meaning is bleached out, giving rise to a deictic meaning, and decategorialization, in that a former noun loses its nominal properties, such as the ability to be inflected and to take modifiers. Finally, there is also erosion in a number of cases, in that the erstwhile noun lost part of its phonetic substance: The Ngiti noun *alɛ* 'person, people' was reduced to *l-* in fast speech as a pronoun but not as a noun, the Saafi-Saafi noun *ɓoʔ* 'person' lost its glottal stop (*ɓoʔ* > *ɓoo*), and the Laala disyllabic noun *ɓoyyi* 'person' was shortened to a monosyllabic form *ɓi*.

3. Grammaticalization in a wide sense

An issue that has figured in a number of discussions on grammatical change is where grammaticalization ends and other phenomena begin, that is, where the limits of grammaticalization are (see e.g. Ramat & Hopper 1998, Bisang, Himmelmann & Wiemer 2004). The observations made in this paper suggest that personal pronouns are the product of regular processes of grammatical

change arising in specific contexts. That these processes are unidirectional is suggested by the following observations:

(a) We have not found any general processes that would contradict the unidirectionality principle. While it is possible that such cases exist, but if they do exist they must be rare.

(b) Many of the processes that we have been dealing with concern a cognitive process leading from a domain that can be defined in terms of physical objects and physical space to the domain of personal deixis, which is defined in terms of the participants of a communicative situation. Accordingly, there appears to be a common conceptual development from referring expressions to address forms – that is, from non-speech participants to speech participants. While it may happen that address forms turn into referential expressions, such a process appears to be clearly less common.

(c) Many of the processes involve relevant parameters of grammaticalization (see (2) in Section 1.2 above and Section 2.3.2.2 above for an example). A linguistic expression is extended to a context suggesting a new meaning, where the literal meaning does not make sense (extension). Once the new meaning is conventionalized, the expression loses the lexical meaning it once had in favor of the new meaning (=desemanticization), and with the transition from lexical to functional meaning it also loses many or all of its morphosyntactic properties (=decategorialization), becoming an invariable marker confined to denoting one specific grammatical function, and at a more advanced stage it may also lose in phonetic substance (=erosion).

(d) The nominal and spatial expressions developing into personal pronouns tend to involve lexical items such as noun phrases or adverb phrases, such as Spanish *Vuestra Merced* 'Your Grace' > *Usted* 'You', while personal pronouns are functional categories, that is, grammatical forms having a schematic meaning. Such processes are therefore in accordance with the general directionality from lexical to grammatical categories. Conversely, we will not expect any directionality whereby pronominal particles turn into morphologically complex structures such as possessive noun phrases or prepositional phrases.

But there are also differences compared to canonical instances of grammaticalization. First, one of the features characterizing grammaticalization is that it tends to lead from referential to non-referential expressions: Nouns and verbs are categories that typically can be referred to anaphorically while this does not apply to functional categories such as tense or case markers (see (3f) above). By contrast, in the development of new personal pronouns, such a change from referential to non-referential function cannot be

observed: Personal pronouns are invariably referential. Accordingly, in order to define the evolution of personal pronouns in terms of grammaticalization, an extended concept of grammaticalization is needed – one that parts with the stipulation that this process leads from referential to non-referential expressions.

Second, and more importantly, none of the processes described in Section 2 can in fact be said to qualify as unambiguous instances of grammaticalization, for the following reason. Grammaticalization can be understood as a process leading to a shift in morphological classes:[34] Linguistic expressions belonging to lexical classes such as those of verbs or nouns change into classes like those of tense-aspect or case markers, etc. Such a shift can in fact be observed in the case of three of the sources that we observed in Section 2, namely those of nouns, spatial deixis, and intensifiers. But obviously, there is no shift in morphological class when plurification (from plural to honorific singular function; see Section 2.2.5 above) or a shift in deixis (from third to second person) are involved. What happens is that a member of the paradigm of personal pronouns changes into another member of the same paradigm. Thus, with the extension of the erstwhile English plural pronoun *you* to also serve as a singular pronoun, *you* continued to be part of the paradigm of English personal pronouns, and the German third person plural pronoun *sie* did not experience any paradigm shift when it acquired the additional function of a second person pronoun (*Sie*) in the 18th century.

And third, we described grammaticalization in Section 1.2 in terms of four parameters. Now, one of these parameters, namely extension, applies to all instances of pronoun genesis. Two other parameters, desemanticization and decategorialization, can be said to be at work when nouns, spatial expressions, and intensifiers develop into personal pronouns, in that the source forms lose in semantic and morphosyntactic properties characterizing their lexical uses. But neither of these parameters is involved when there is plurification or shift in deixis.

Table 4 (on next page) provides an overview of the characteristics associated with the evolution of personal pronouns. In the case of canonical instances of grammaticalization one would expect that all criteria distinguished be fulfilled. Accordingly, there is reason to argue that none of the cases discussed in this paper qualifies as an instance of grammaticalization. Alternatively, we argue that we are dealing with an extended notion of grammaticalization since two core properties are present, namely extension and – most importantly – unidirectionality. We propose to call the latter GRAMMATICALIZATION IN A WIDE SENSE (GWS) to distinguish it from the

[34] This shift is gradual rather than discontinous (see the model of extension in Table 5 below), that is, there are intermediate stages where e.g. an item such as a verb can have both lexical (verbal) and functional (tense/aspect) properties.

Source	Unidirect-ionality	Paradigm shift	Extension	Desemant-icization	Decategor-ialization	Change in referentiality
Nouns	+	+	+	+	+	−
Spatial deixis	+	+	+	+	+	−
Intensifiers	+	+	+	+	+	−
Deicitc shift	+	(+)	+	(+)	(+)	−
Plurification	+	−	+	+	−	−

Table 4
The evolution of personal pronouns and parameters of grammaticalization. (In this table, we are ignoring the parameter of erosion since it may, but need not, accompany grammaticalization; see Section 1.2.)

former, which is in accordance with all the criteria listed in Table 4 and may be referred to as GRAMMATICALIZATION IN A NARROW SENSE (GNS).

What the notion GWS exactly means with reference to a more general understanding of grammaticalization is an issue that is beyond the scope of the present paper and requires much further research. Such research needs to take other, and more general concepts into account, like that of 'super-grammaticalization' as proposed by Lindström Tiedemann (2006), which goes one step further than the present proposal by also subsuming phenomena such as word order, intonation, and sound change as showing manifestations of grammaticalization.

3.1 *The extension model*

A number of approaches have been used to deal with the phenomena relating to extension (see e.g. Bybee et al. 1994; Traugott & Dasher 2002: 34–39); in the present paper we use the pragmatic model of context-induced reinterpretation of Heine (2002), presented in Table 5 (on next page), to describe the most salient characteristics of extension.[35]

Table 5 suggests that the transition from a meaning of stage I to one of stage IV does not proceed straight from one to another; rather, it involves two intermediate stages, namely stages II and III. The model of extension has been shown to be applicable to canonical instances of grammaticalization, and it also accounts for the development of personal pronouns as described in the preceding sections.

We may illustrate the relevance of the model by looking again at the evolution from generic human noun ('people' or 'person(s)') to first person plural pronouns that we discussed in Section 2.3.2.1. The situation in the Labwor language is characteristic of the bridging stage II, where the noun *jɔ* 'people' expresses in certain contexts the meaning of a first person plural pronoun, even if the meaning 'people' is never entirely ruled out. This innovation is confined to very few contexts, and not every Labwor speaker makes use of it. Furthermore, it shows a low frequency of occurrence, its use is optional, and speakers do not to consider it to be a distinct new category; rather, it tends to be interpreted as a deviant application of the old category of nouns.

While the bridging context of stage II is suggestive of the incipient stage of grammaticalization, the switch stage III leads to the establishment of a new grammatical category. The situation of *a gente* in European Portuguese is an instance of this stage. In certain contexts the nominal reading ('the people') makes little, if any, sense; rather, *a gente* can only be understood as a category of personal deixis ('we') in such contexts. Furthermore, it is used in its new function fairly frequently and in a wider range of social and linguistic

[35] See Heine (2002) for more details; see also Diewald (2002) for a similar model.

Stage	Context	Resulting meaning	Type of inference
I Initial	Unconstrained	Source meaning	—
II Bridging context	There is a new context triggering a new meaning	Target meaning foregrounded	Invited (cancellable)
III Switch context	There is a new context which is incompatible with the source meaning	Source meaning backgrounded	Usual (typically non-cancellable)
IV Conventionalization	The target meaning no longer needs to be supported by the context that gave rise to it; it may be used in new contexts	Target meaning only	—

Table 5
A model of extension (context-induced reinterpretation).

contexts. At the same time, however, there remain contexts where its use is barred, i.e. social and linguistic contexts that remain reserved for the old category of a noun, and while used quite regularly, speakers need not use it in all environments where it might be expected.

It is only at the final, conventionalization stage IV that the innovation has attained the status of a new full-fledged grammatical category, where it no longer shows contextual restrictions, being used roughly as frequently as other members of the same paradigm, and showing the same degree of obligatoriness as the other members of that paradigm. In addition, the innovated category may lose morphosyntactic properties that were characteristic of its former category status but are no longer in line with the morphosyntax of the new paradigm. Being a stage III category in European Portuguese, *a gente* is approaching a stage IV situation in Brazilian Portuguese (Albán & Freitas 1991; Schwegler 1993: 152; Merlan 2006: 235; Travis & Silveira 2009; Martelotta & Cezario in press): It has become a regular member of a newly emerging paradigm of personal pronouns and, accordingly, is used generally in this new function (see Section 2.3.2.1 above). The fact that *a gente* has largely replaced the earlier pronoun *nós* 'we'[36] in Brazilian Portuguese suggests that its use has been generalized to the extent that it is employed in contexts that formerly were reserved for *nós*, hence that it is largely conventionalized (stage IV). While *gente* continues to exist as a lexical noun ('people'), the noun *gente* and the pronoun *a gente* now appear to be distinct categories,[37] with the relationship between the two being more appropriately described as one of homonymy rather than of polysemy.[38]

Table 6 (on next page) summarizes the main pragmatic features associated with the unidirectional evolution of both kinds of grammaticalization. It would seem that all developments leading to new expressions for personal deixis that we discussed in this paper are in accordance with the extension model (Table 5), but some lack salient properties of GNS. Plurification in particular lacks what in Table 4 above is referred to as paradigm shift, such as a shift from open to closed class category or from one kind of closed class category to another: When a plural pronoun is extended to also serve for honorific singular address then it does not shift from one paradigm to

[36] Contexts where *nós* has survived include the frequently used verbs *ter* 'have' and *ser* 'be', particularly in the present indicative, e.g. *nós temos* 'we have', *nós somos* 'we are' (Travis & Silveira 2009: 349).

[37] For example, with the grammaticalization of English *be going to* to a future marker (e.g. *Paul is going to come*), the corresponding lexical use of *be going to* did not disappear (*Paul is going to London*). This co-existence of a grammaticalized item with its non-grammaticalized source item is described in works on grammaticalization as 'divergence' (Hopper 1991).

[38] Lichtenberk (1991) proposes the term 'heterosemy' for cases such as *a gente* of Brazilian Portuguese.

Stage	Context	Frequency	Use characteristics
II Bridging context	Highly restricted	Low	Optional
III Switch context	Larger range of contexts	High	Regular
IV Conventionalization	Generalized	Very high	Obligatory

Table 6
Some pragmatic characteristics of extension in grammaticalization.

another; rather, it remains a member of the same paradigm of person markers, hence it does not qualify as an instance of grammaticalization in the narrow sense (GNS). We propose to call this kind of grammatical change, which takes place within one and the same morphosyntactic paradigm, a LATERAL SHIFT, a term proposed by Joseph (2005).[39] We argue that lateral shift falls squarely within the domain of grammaticalization in a wide sense (GWS) in that it conforms to both the unidirectionality principle and the model of extension.[40] The exact boundary between GNS and GWS is unclear; there are types of grammatical change that can be classified as instances of both. More research is needed on this issue.

4. CONCLUSIONS

In a survey of the kind carried out here it is not possible to draw on a representative sample of languages; what we thought we could achieve was simply to find as much information from as many languages as possible in order to reconstruct crosslinguistically regular mechanisms leading to the creation of new personal pronouns. Accordingly, the languages figuring in this paper are neither areally nor genetically evenly distributed; certain regions of the world, such as Australia or the Americas, are clearly

[39] 'Lateral shifts can be defined as a change in the form of a grammatical affix that is not just a simple sound change (and so is a "higher level", grammatical, change) but does not alter the element's grammatical nature or status in terms of where it falls on the "cline" of grammatical status (from word to affix). Thus after the change, the element in question is neither more nor less grammatical than before.' (Joseph 2005: 1–2).

[40] For another case of purported lateral shift, see Luraghi (2005: 16), who argues that prepositions constitute a fuzzy domain in that 'we have the same type of linguistic unit (a preposition) before and after the change, the preposition has not become more obligatory, and it has not taken a more grammatical function, it has simply extended its meaning'.

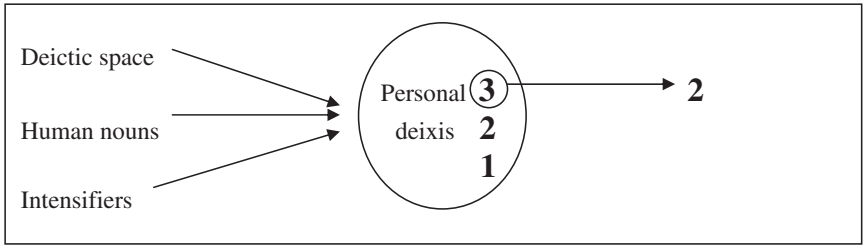

Figure 1
The main conceptual sources and pathways leading to personal pronouns.

underrepresented while others, in particular Southeast and East Asia, Europe, and Africa, are more strongly represented than might be justified. What effect this has on the results obtained is a question that is open to future research.

The preceding discussions may have shown that personal deixis is not necessarily a crosslinguistically neat and stable conceptual domain. We observed in Section 2 that in some languages of Southeast and East Asia there is some pragmatically defined fluidity in the deictic significance of grammatical forms: What serves as a second person pronoun in one context can be used as a first person pronoun or an intensifier (reflexive) in another context. We pointed out, for example, that since the 16th century the Korean form *jeo* was used as an intensifier/reflexive form for all honorific levels, later on being degraded to a low level form, and around the end of the 19th century, *jeo* became a first person pronoun of the low honorific level. And Khmer does not really have a second person pronoun, although the intensifier '*aeng* 'oneself' is commonly used in this function ('you (plural)'). However, if there is a close personal relationship between speaker and hearer, then '*aeng* may as well refer to the first person ('I' or 'we'; John Haiman, p.c.).

We saw in Section 2 that there is a limited pool of conceptual sources that speakers tend to recruit in order to create new forms of personal pronouns, most of all the sources listed in (7) above. Using the emergence of the Calunga language in Brazil as an example, Heine & Song (2010; see also Byrd 2006) demonstrate how an entirely new system of personal pronouns can arise. This creation was based essentially on the exploitation of two conceptual sources, namely nominal concepts and spatial deixis.

Figure 1 rests on generalizations on the conceptual domains that appear to provide the major cognitive templates for the development of personal pronouns.[41] Both this figure and generally the discussion in this paper are based

[41] Note that we have ignored the plurification strategy in both presentations.

on a number of simplifications of the phenomena examined. These phenomena are the product of a multitude of communal acts of language use extending over long periods of time. But what ends up as a communal act is likely to have started as an individual act, where a certain rhetorical strategy was propagated by someone, adopted by others, in some cases spreading across the entire speech community, thereby leading to a new form of expressing personal deixis. This individual act is no longer recoverable; the findings presented here therefore have to be taken with care: They are based on hypotheses on how people interacted in the past, rather than on 'facts' that are readily accessible to the student of language use. The general motivation underlying the pathways distinguished in Figure 1 can be seen in human activity aimed at finding optimal ways of saying what is both socially appropriate and most advantageous for the speaker in a given sociolinguistic context.

REFERENCES

Albán, Maria del Rosário & Judith Freitas. 1991. *Nós* ou *gente*? *Estudos* 11, 75–89.

Arends, Jacques. 1986. Genesis and development of the equative copula in Sranan. In Pieter Muysken & Norval Smith (eds.), *Substrata versus universals in creole genesis* (Creole Language Library 1), 103–127. Amsterdam & Philadelphia: John Benjamins.

Baker, Philip. 1995. Some developmental inferences from the historical studies of pidgins and creoles. In Jacques Arends (ed.), *The early stages of creolization* (Creole Language Library 13), 1–24. Amsterdam & Philadelphia: John Benjamins.

Barth-Weingarten, Dagmar & Elizabeth Couper-Kuhlen. 2002. On the development of final *though*: A case of grammaticalization? In Wischer & Diewald (eds.), 345–361.

Bisang, Walter, Nikolaus P. Himmelmann & Björn Wiemer (eds.). 2004. *What makes grammaticalization? A look from its fringes and its components* (Trends in Linguistics, Studies and Monographs 158). Berlin: Mouton de Gruyter.

Blake, Frank R. 1934. The origin of pronouns of the first and second persons. *The American Journal of Philology* 55, 244–248.

Brown, Roger & Albert Gilman. 1968. The pronouns of power and solidarity. In Joshua A. Fishman (ed.), *Readings in the sociology of language*, 252–281. The Hague & Paris: Mouton.

Bybee, Joan L., Revere D. Perkins & William Pagliuca. 1994. *The evolution of grammar: Tense, aspect and modality in the languages of the world*. Chicago: University of Chicago Press.

Byrd, Steven Eric. 2006. *Calunga, an Afro-Brazilian speech of the Triangulo Mineiro*. Ann Arbor, MI: University of Michigan Services.

Campbell, Lyle & Richard Janda. 2001. Introduction: Conceptions of grammaticalization and their problems. *Language Sciences* 23.2–3, 93–112.

Casad, Eugene 1984. Cora. In Ronald W. Langacker (ed.), *Studies in Uto-Aztecan grammar*, vol. 4: *Southern Uto-Aztecan grammatical sketches* (Summer Institute of Linguistics Publications in Linguistics 56, IV), 151–459. Arlington, TX: The Summer Institute of Linguistics & The University of Texas at Arlington.

Claudi, Ulrike 1985. *Zur Entstehung von Genussystemen: Überlegungen zu einigen theoretischen Aspekten, verbunden mit einer Fallstudie des Zande*. Hamburg: Helmut Buske.

Comrie, Bernard. 1975. Polite plurals and predicate agreement. *Language* 51, 406–441.

Cooke, Joseph Robinson. 1968. *Pronominal reference in Thai, Burmese, and Vietnamese* (University of California Publications in Linguistics 52). Berkeley, CA: University of California Press.

Corne, Chris. 1971. Le patois créole français de la Guyane (St-Laurent-du-Maroni): esquisse de grammaire. *Te Reo* 14, 81–103.
Coulmas, Florian. 1992. Linguistic etiquette in Japanese society. In Richard J. Watts, Sachiko Ide & Konrad Ehlich (eds.), *Politeness in language: Studies in its history, theory and practice* (Trends in Linguistics, Studies and Monographs 59), 299–323. Berlin & New York: Mouton de Gruyter.
Coveney, Aidan. 2000. Vestiges of *nous* and the 1st person plural verb in informal spoken French. *Language Sciences* 22, 447–481.
Croft, William. 2000. *Explaining language change: An evolutionary approach*. London: Longman.
Cysouw, Michael. 2003. *The paradigmatic structure of person marking* (Oxford Studies in Typology and Linguistic Theory). Oxford: Oxford University Press.
Cysouw, Michael. 2005a. Inclusive/exclusive distinction in independent pronouns. In Haspelmath et al. (eds.), 162–164.
Cysouw, Michael. 2005b. Inclusive/exclusive distinction in verbal inflection. In Haspelmath et al. (eds.), 164–169.
Diessel, Holger. 1997. The diachronic reanalysis of demonstratives in crosslinguistic perspective. *Chicago Linguistic Society* (*CLS*) 33, 83–98.
Diessel, Holger. 1999a. The morphosyntax of demonstratives in synchrony and diachrony. *Linguistic Typology* 3, 1–49.
Diessel, Holger. 1999b. *Demonstratives: Form, function, and grammaticalization* (Typological Studies in Language 42). Amsterdam & Philadelphia: John Benjamins.
Diewald, Gabriele. 2002. A model for relevant types of contexts in grammaticalization. In Wischer & Diewald (eds.), 103–120.
Forchheimer, Paul. 1953. *The category of person in language*. Berlin: Walter de Gruyter.
Friedländer, Marianne. 1974. *Lehrbuch des Susu*. Leipzig: Enzyklopädie.
Gardiner, A. 1957. *Egyptian grammar: Being an introduction to the study of hieroglyphs*, 3rd edn. Oxford: Oxford University Press.
Givón, Talmy. 1984. *Syntax: A functional–typological introduction*, vol. 1. Amsterdam & Philadelphia: John Benjamins.
Gottsched, Johann Christoph. [1762] 1970. *Vollständigere und Neuerläuterte Deutsche Sprachkunst*, reprint. Hildesheim & New York.
Greenberg, Joseph H. (ed.). 1963a. *Universals of language*. Cambridge, MA: MIT Press.
Greenberg, Joseph H. 1963b. Some universals of grammar with particular reference to the order of meaningful elements. In Greenberg (ed.), 58–90.
Günthner, Susanne & Katrin Mutz. 2004. Grammaticalization vs. pragmaticalization? The development of pragmatic markers in German and Italian. In Bisang et al. (eds.), 77–107.
Haase, Martin. 1992. *Sprachkontakt und Sprachwandel im Baskenland: die Einflüsse des Gaskognischen und Französischen auf das Baskische*. Hamburg: Buske.
Haase, Martin. 1994. *Respekt: die Grammatikalisierung von Höflichkeit* (Edition Linguistik 44). Munich & Newcastle: LINCOM Europa.
Hagège, Claude. 1993. *The language builder: An essay on the human signature in linguistic morphogenesis* (Amsterdam Studies in the Theory and History of Linguistic Science 94). Amsterdam & Philadelphia: John Benjamins.
Hartmann, Josef. 1980. *Amharische Grammatik* (Äthiopische Forschungen 3). Wiesbaden: Franz Steiner.
Haspelmath, Martin, Matthew S. Dryer, David Gil & Bernard Comrie (eds.). 2005. *The world atlas of language structures*. Oxford: Oxford University Press.
Head, Brian. 1978. Respect degrees in pronominal reference. In Joseph H. Greenberg, Charles A. Ferguson & Edith Moravcsik (eds.), *Universals of human language*, vol. 3: *Word structure*, 151–211. Stanford, CA: Stanford University Press.
Heath, Jeffrey. 2004. Person. In Geert E. Booij, Christian Lehmann, Joachim Mugdan & Stavros Skopeteas (eds.), *Morphologie – Morphology: Ein internationales Handbuch zur Flexion und Wortbildung* (Handbücher zur Sprach- und Kommunikationswissenschaft 17.2), vol. 2, 998–1015. Berlin & New York: Walter de Gruyter.
Heine, Bernd. 1999. *The ǁAni: Grammatical notes and texts* (Khoisan Forum 11). Cologne: Institut für Afrikanistik.
Heine, Bernd. 2002. On the role of context in grammaticalization. In Wischer & Diewald (eds.), 83–101.

Heine, Bernd, Ulrike Claudi & Friederike Hünnemeyer. 1991. *Grammaticalization: A conceptual framework*. Chicago: University of Chicago Press.
Heine, Bernd & Tania Kuteva. 2002. *World lexicon of grammaticalization*. Cambridge: Cambridge University Press.
Heine, Bernd & Mechthild Reh. 1984. *Grammaticalization and reanalysis in African languages*. Hamburg: Buske.
Heine, Bernd & Kyung-An Song. 2010. On the genesis of personal pronouns: Some conceptual sources. *Language and Cognition* 2.1, 117–148.
Helmbrecht, Johannes. 2004. *Personal pronoun: Form, function and grammaticalization* (Diversitas Linguarum, Universität Erfurt). Bochum: Universitätsverlag Brockmeyer.
Helmbrecht, Johannes. 2005a. Typologie und Diffusion von Höflichkeitspronomina in Europa. *Folia Linguistica* 39.3–4, 417–452.
Helmbrecht, Johannes. 2005b. Politeness distinctions in pronouns. In Haspelmath et al. (eds.), 186–189.
Hengeveld, Kees. In press. The grammaticalization of tense and aspect. In Narrog & Heine (eds.).
Heusing, Gerald. 2004. *Die südlichen Lwoo-Sprachen: Beschreibung, Vergleich und Rekonstruktion* (Nilo-Saharan 19). Cologne: Köppe.
Hinds, John. 1986. *Japanese*. London & New York: Routledge.
Hopper, Paul J. 1991. On some principles of grammaticization. In Traugott & Heine (eds.), 17–35.
Hopper, Paul J. & Elizabeth C. Traugott. 2003. *Grammaticalization*. Cambridge: Cambridge University Press.
Joseph, Brian D. 2005. How accommodating of change is grammaticalization? The case of 'lateral shifts'. *Logos and Language* 6.2, 1–8.
Keesing, Roger M. 1991. Substrates, calquing and grammaticalization in Melanesian Pidgin. In Traugott & Heine (eds.), 315–342.
Klausenburger, Jürgen. 2000. *Grammaticalization: Studies in Latin and Romance morphosyntax* (Amsterdam Studies in the Theory and History of Linguistic Science 193). Amsterdam & Philadelphia: John Benjamins.
König, Ekkehard & Peter Siemund. 2000. Intensifiers and reflexives: A typological perspective. In Zygmunt Frajzyngier & Traci S. Curl (eds.), *Reflexives: Forms and functions* (Typological Studies in Language 40), 41–74. Amsterdam & Philadelphia: John Benjamins.
Kuteva, Tania. 2001. *Auxiliation: An enquiry into the nature of grammaticalization*. Oxford & New York: Oxford University Press.
Kutsch Lojenga, Constance. 1994. *Ngiti: A Central-Sudanic language of Zaire* (Nilo-Saharan 9). Cologne: Köppe.
Lakoff, George. 1987. *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago: University of Chicago Press.
Lakoff, George & Mark Johnson. 1980. *Metaphors we live by*. Chicago: University of Chicago Press.
Lehmann, Christian. 1985. Grammaticalization: Synchronic variation and diachronic change. *Lingua e Stile* 20 303–318.
Lehmann, Christian. 1995. *Thoughts on grammaticalization*. Munich: LINCOM Europa.
Levine, Robert D. 2010. The *ass* camouflage construction: Masks as parasitic heads. *Language* 86.2, 265–301.
Lewis, Geoffrey L. 1985. *Turkish grammar*. Oxford: Oxford University Press.
Lichtenberk, Frantisek. 1991. Semantic change and heterosemy in grammaticalization. *Language* 67.3, 475–509.
Lindström Tiedemann, Therese. 2006. Grammaticalization – past and present. *Logos and Language* 6.2, 19–35.
Lockwood, W. B. 1968. *Historical German syntax*. Oxford: Clarendon.
Luraghi, Silvia. 2005. Does a theory of language change need unidirectionality? *Logos and Language* 6.2, 9–17.
Margetts, Anna & Peter K. Austin. 2007. Three-participant events in the languages of the world: Towards a crosslinguistic typology. *Linguistics* 45.3, 393–451.
Martelotta, Mário Eduardo & Maria Maura Cezario. In press. Grammaticalization in Brazilian Portuguese. In Narrog & Heine (eds.).

Merlan, Aurelia. 2006. Grammatikalisierungstendenzen im Portugiesischen und Rumänischen: von Nominalsyntagmen zu Pronomina. In Jürgen Schmidt-Radefeldt (ed.), *Portugiesisch kontrastiv gesehen und Anglizismen weltweit* (Rostocker Romanistische Arbeiten 10), 221–240. Frankfurt: Peter Lang.
Mühlhäusler, Peter & Rom Harré. 1990. *Pronouns and people: The linguistic construction of social and personal identity* (Language in Society 15). Oxford: Blackwell.
Muysken, Pieter & Norval Smith (eds.). 1986. *Substrata versus universals in Creole genesis: Papers from the Amsterdam Creole Workshop, April 1985*. Amsterdam & Philadelphia: John Benjamins.
Narrog, Heiko & Bernd Heine (eds.). In press. *The Oxford handbook of grammaticalization*. Oxford: Oxford University Press.
Norde, Muriel. In press. Degrammaticalization. In Narrog & Heine (eds.).
Postal, Paul. 1969. On so-called pronouns in English. In David Reibel & Sanford Schane (eds.), *Modern studies in English*, 187–224. Englewood Cliffs, NJ: Prentice Hall.
Praetorius, Franz. 1879. *Die amharische Sprache*. Halle: Verlag der Buchhandlung des Waisenhauses.
Ramat, Anna Giacalone & Paul J. Hopper (eds.). 1998. *The limits of grammaticalization*. Amsterdam & Philadelphia: John Benjamins.
Samarin, William J. 2002. Plurality and deference in urban Sango. In Vicki Carstens & Frederick Parkinson (eds.), *Advances in African linguistics* (Trends in African Linguistics 4), 299–311. Trenton, NJ & Asmara: Africa World Press.
Schwegler, Armin. 1993. Subject pronouns and person/number in Palenquero. In Francis X. Byrne & John A. Holm (eds.), *Atlantic meets Pacific: A global view of pidginization and creolization* (Creole Language Library 11), 145–161. Amsterdam & Philadelphia: John Benjamins.
Shibatani, Masayoshi. 1990. *The languages of Japan*. Cambridge: Cambridge University Press.
Siewierska, Anna. 2004. *Person*. Cambridge: Cambridge University Press.
Simon, Horst J. 1997. Die Diachronie der deutschen Anredepronomina aus Sicht der Universalienforschung. *Sprachtypologie und Universalienforschung* 50, 267–281.
Simon, Horst J. 2003. *Für eine Grammatische Kategorie 'Respekt' im Deutschen: Synchronie, Diachronie und Typologie der deutschen Anredepronomina*. Tübingen: Max Niemeyer.
Sohn, Ho-Min. 1999. *The Korean language*. Cambridge: Cambridge University Press.
Song, Kyung-An. 2002. Korean reflexives and grammaticalization: A speaker–hearer dynamic approach. *Sprachtypologie und Universalienforschung* (*STUF*, Berlin) 55.5, 340–353.
Song, Kyung-An. 2010. Pronominals, addressing forms, and grammaticalization: A speaker–hearer dynamic approach with reference to Korean. Ms., Chonnam National University, Gwangju.
Sugamoto, Nobuko. 1989. Pronominality: A noun–pronoun continuum. In Roberta Corrigan, Fred Eckman & Michael Noonan (eds.), *Linguistic categorization*, 267–291. Amsterdam & Philadelphia: John Benjamins.
Syea, Anand. 1996. The development of a marker of definiteness in Mauritian Creole. In Philip Baker & Anand Syea (eds.), *Changing meanings, changing functions: Papers relating to grammaticalization in contact languages* (Westminster Creolistics 2), 171–186. London: University of Westminster Press.
Traugott, Elizabeth C. & Richard B. Dasher. 2002. *Regularity in semantic change* (Cambridge Studies in Linguistics 96). Cambridge: Cambridge University Press.
Traugott, Elizabeth C. & Bernd Heine (eds.). 1991a. *Approaches to grammaticalization*, vol. 1. Amsterdam & Philadelphia: John Benjamins.
Traugott, Elizabeth C. & Bernd Heine (eds.). 1991b. *Approaches to grammaticalization*, vol. 2. Amsterdam & Philadelphia: John Benjamins.
Travis, Catherine E. & Agripino S. Silveira. 2009. The role of frequency in first-person plural variation in Brazilian Portuguese: *Nós* vs. *a gente*. *Studies in Hispanic and Lusophone Linguistics* 2.2, 347–376.
Tucker, Archibald N. 1940. *The Eastern Sudanic languages*, vol. 1. London, New York & Toronto: Oxford University Press.
Wischer, Ilse & Gabriele Diewald (eds.). 2002. *New reflections on grammaticalization* (Typological Studies in Language 49). Amsterdam & Philadelphia: John Benjamins.

Wordick, F. J. F. 1982. *The Yindjibarndi language* (Pacific Linguistics C-71). Canberra: The Australian National University, Research School of Pacific Studies.

Authors' addresses: (*Heine*)
Institut für Afrikanistik, Universität zu Köln, 50923 Köln, Germany
Correspondence address: Bernd Heine, Nonnenwerthstr. 48, 50937, Koeln, Germany
bernd.heine@uni-koeln.de

(*Song*)
Department of German Language & Literature, Chonnam National University, 77 Yongbong-ro, Buk-gu, Gwangju 500-757, Korea
kasong@chonnam.ac.kr