

Double Double, Morphology and Trouble: Looking into Reduplication in Indonesian

Meladel Mistica, Avery Andrews, I Wayan Arka

The Australian National University
{meladel.mistica,avery.andrews,
wayan.arka}@anu.edu.au

Timothy Baldwin

The University of Melbourne
tb@ldwin.net

Abstract

This paper investigates reduplication in Indonesian. In particular, we focus on verb reduplication that has the agentive voice affix *meN*, exhibiting a homorganic nasal. We outline the recent changes we have made to the implementation of our Indonesian grammar, and the motivation for such changes.

There are two main issues that we deal with in our implementation: how we account for the morphophonemic facts relating to sound changes in the morpheme; and how we construct word formation (i.e. sublexical) rules in creating these derived words exhibiting reduplication.

1 Introduction

This study looks at full reduplication in Indonesian verbs, which is a morphological operation that involves the doubling of a lexical stem. In this paper, we step through the word formation process of reduplication involving agentive voice marking, including the morphophonemic changes and the morphosyntactic changes brought about by this construction. The reduplication investigated here is a productive morphological process; it is readily applied to many lexical stems in creating new words. Instead of having extra entries in the lexicon for reduplicated words, we aim to investigate the changes brought about by reduplication and encode them in a meaningful way to interpret, during parsing, these morphosyntactically complex, valance-changing, derived words.

This investigation sits within a larger Indonesian resource project that primarily aims to build an elec-

tronic grammar for Indonesian within the framework of *Lexical Functional Grammar* (LFG). Our project forms part of a group of researchers, PARGRAM¹ whose aim is to also produce wide-coverage grammars built on a collaboratively agreed upon set of grammatical features (Butt et al., 1999). In order to ensure comparability we use the same linguistic tools for implementation.²

One of the issues we address is how to adequately account for morphophonemic facts, as schematised in Examples (1), (2) and (3):

- (1) $[meN+tarik]^2$
 $\leftrightarrow meN+tarik+hyphen+meN+tarik$
 $\leftrightarrow menarik-menarik$
“pulling (iteratively)”
- (2) $meN+[tarik]^2$
 $\leftrightarrow meN+tarik+hyphen+tarik$
 $\leftrightarrow menarik-narik$ (**menarik-tarik*)
“pulling quickly”
- (3) $meN+[tarik]^2$
 $\leftrightarrow tarik+meN+hyphen+tarik$
 $\leftrightarrow tarik-menarik$ (**narik-menarik*)
“pull at each other”

Here, *tarik* “pull” is the verb stem, *meN* is a verbal affix with a homorganic nasal (the function of which will be discussed in Section 2.1), 2 is the notation we use for reduplication, and the square brackets [] are used to specify the scope of the reduplication.

¹<http://www2.parc.com/is1/groups/nl/tt/pargram/>

²<http://www2.parc.com/is1/groups/nl/tt/xle/> and <http://www.stanford.edu/~laurik/fsmbook/home.html>

Each of the examples consists of three lines: (a) a simplified representation of which words are reduplicated, (b) a breakdown of the components that make up the surface word, and (c) the surface word (in italics). Note that the first-line representation for (2) and (3) is identical, but the surface words differ on the basis of the order in which the reduplication and *meN* affixation are applied. Note also that, as is apparent in the gloss, (3) involves a different process to the other two examples, and yet all three are dealt with using the same reduplication strategy in our implementation. We return to discuss these and other issues in Section 3.

The morphological analyser is based on the system built by Pisceldo et al. (2008), whose implementation of reduplication follows closely that suggested for Malay by Beesley and Karttunen (2003). However, (3) is not dealt with by Beesley and Karttunen (2003), and the solution of Pisceldo et al. (2008) requires an overlay of corrections to account for the distinct argument structure of (3). This paper outlines a method for reorganising the morphological analyser to account for these facts in a manner which is more elegant and faithful to the data.

2 Reduplication in Indonesian

2.1 About Indonesian

Indonesian is a Western Austronesian language that has **voice marking**, which is realised as an affix on the verb that signals the thematic status of the subject (Musgrave, 2008). In Indonesian, the subject is the left-most NP in the clause. Below we see examples of AV (agentive voice),³ PV (patient or passive voice) and UV (undergoer voice — bare stem).

- (4) [*Amir*] *membaca* *buku* *itu*
Amir AV+read book this
“Amir read the book”
- (5) [*Buku itu*] *dibaca* *oleh* *Amir*
book this PV+read by Amir
“The book was read by Amir”
- (6) [*Temannya*] *dia* *pukul*
his.friend he/she UV.hit
“He hit his friend”

³In (4) the *mem-* “AV- AGENTIVE VOICE” is actually *me* plus a homorganic nasal

The marking on the verb indicates the semantic role of the subject, in square braces [] the agent in (4), and the theme and patient in (5) and (6).

2.2 Productive Reduplication

Indonesian has three types of reduplication: partial, imitative and full reduplication (Sneddon, 1996). We only consider full reduplication — or full repeat of the lexical stem — for this study because it is the only type of reduplication that is productive. We encode three kinds of full reduplication in the morphological analyser:

(7) REDUPLICATION OF STEM

<i>duduk-duduk</i>	<i>sakit-sakit</i>
sit-sit	sick-sick
“sit around”	“be periodically sick”

(8) REDUPLICATION OF STEM WITH AFFIXES

<i>membunuh-bunuh</i>	<i>bunuh-membunuh</i>
AV+hit-hit	AV+hit-hit
“hitting”	“hit each other”

(9) REDUPLICATION OF AFFIXED STEM

<i>membeli-membeli</i>
AV+buy-AV+buy
“buying”

Reduplication seems to perform a number of different operations. There is an aspectual operation, which affects how the action is performed over time. These examples are seen in (7) *sakit-sakit* and (8) *membunuh-bunuh*. These are comparable to the English progressive *-ing* in *He is kissing the vampire* versus *He kissed the vampire*, where the former depicts an event performed over time and the latter a punctual one.

However, this operation is not exactly equivalent to the English progressive, as seen below:

- (10) *Saya memukul-memukul dia*
1.SG AV+hit-AV+hit 3.SG
“I am/was hitting him”/“I repeatedly hit him.”
- (11) #*Saya membunuh-membunuh dia*
1.SG AV+kill-AV+kill 3.SG
“#I was killing him”

(12) *Saya membunuh binatang*
 1.SG AV+hit-AV+hit animal
 “I killed an animal”/“I killed animals”

(13) *Saya membunuh-membunuh binatang*
 1.SG AV+hit-AV+hit animal
 “I killed animal after animal”/“#I was killing the animal”

As can be seen, this operation cannot apply to the verb *bunuh* “kill” in (11) to mean “killing”. However if the object can be interpreted as plural then the action can be applied to the multiple objects as shown in (13). So there is this sense of either being able to distribute the action over time repeatedly or distribute/apply the action over different objects, when the semantics of the event does not allow the action to be repeated again and again, such as killing one animal.⁴ The examples in (7) show more semantic variation on reduplication, such as an additional meaning of purposelessness for *duduk-duduk* “sit around”.⁵

Another function of reduplication is the formation of reciprocals, as shown with *bunuh-membunuh* in (8). This verb formation is clearly not simply a case of reduplicating an affixed stem; there is a more involved process. We see that this kind of reduplication involves valence reduction: in (14) we have a subject and an object that’s expressed in the sentence, but in (15) we only have a subject expressed, which encodes both the agent and patient.

(14) *Mereka membunuh dia.*
 they AV+kill him/her
 “They kill him/her.”

(15) *Mereka bunuh-membunuh*
 they kill-AV+kill
 “They kill each other.”

3 Tools to Construct the Word

This section outlines the process for building up the word. We look at the tools that are used and the theoretical framework upon which the tools are built.

⁴The example in (11) can only be felicitously used if your victim was part of the army of the undead - FYI.

⁵These types of examples will not be discussed further here as they do not exhibit agentive voice marking.

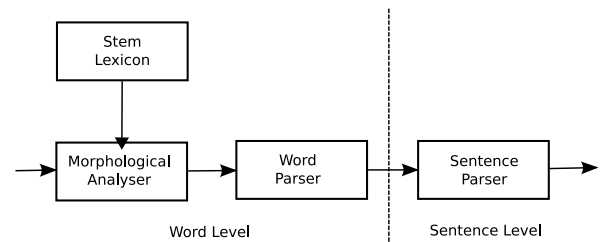


Figure 1: Pipeline showing word-level and sentence-level processes

Figure 1 is the overall course-grained architecture of the system. The dotted vertical line in Figure 1 delimits the boundary between sublexical processes and sentential (or partial) parsing. We are only interested in discussing the components to the left of this boundary, which is where the building of the word-level processes take place.

The components marked “Stem Lexicon” and “Morphological Analyser” utilise the finite state tools XFST and LEXC. The input to the morphological analyser is the sentence that has been tokenised, and its output is a representation of the words split into its morphemes. Furthermore, the first lines of each of the examples of (1), (2), and (3) seen earlier are the representation used, but simplified here to show only the required detail; they show the parts of the word are reduplicated and what other affixes are exhibited. This is then fed as input to the “Word Parser”.

3.1 Theoretical Assumptions

The grammar formalism upon which the ‘Word Parser’ and ‘Sentence Parser’ are built is Lexical Functional Grammar (LFG). LFG has ‘a parallel correspondence’ architecture (Bresnan, 2001), which means relevant syntactic information is distributed among the parallel representations, and that the representations are related via mapping rules.

The level of representation that defines grammatical functions (subject, object etc.) and the constraints upon them, as well as features such as tense and aspect is called the **f-structure**. The f-structure is represented as attribute value matrices, where all required attributes must have unique and complete values. The **c-structure** is represented with phrase

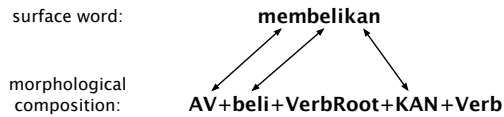


Figure 2: Upper and lower language correspondence for *membelikan* “buy someone something”

structure trees and describes the language-specific arrangement of phrases and clauses for a given language. This level of representation accounts for the surface realisation of sentences, such as word order. The **a-structure** specifies the arity of the predicate, defining its arguments and their relative semantic prominence, which have mapping correspondences to grammatical functions.

3.2 Finite State Tools: XFST and LEXC

The ‘Morphological Analyser’ is built with tools that provide access to finite-state calculus algorithms, in particular the XEROX FINITE-STATE CALCULUS implementation (Beesley and Karttunen, 2003). The finite-state network we create with these tools is a transducer, which allows for a lower language — or a definition of the allowable surface words in the language — and an upper language, which defines the linear representation of the morphological units in the surface word. An example of an upper language ‘output’, for analysis, and its corresponding lower language ‘input’ or morphological analysis is given in Figure 2.

In this example the *mem-* prefix is represented with AV+, the stem *beli* “buy” gets extra information about its part-of-speech via the +VerbRoot suffix, and the applicative *-kan* is represented as +KAN.

We encode the morphotactics of the Indonesian word with the XFST tool, which provides an interface to these algorithms for defining and manipulating the finite state networks, as well as LEXC, which is used for defining the lexicon (Beesley and Karttunen, 2003).

The Pisceldo et al. (2008) system, on which our system is based, employs the same finite state tools as the current implementation. It has two major components which are labelled **morphotactic rules** and **morphophonemic rules**. Figure 3 shows the general schema of the Pisceldo et al. (2008) system.

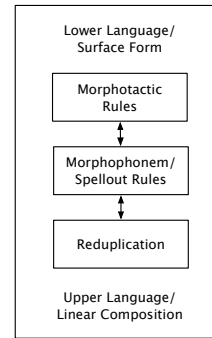


Figure 3: Pisceldo et al. (2008) morphological analyser

The label **reduplication** is a little misleading because it simply indicates when the doubling of the morphological form takes place. In XFST this process is named **compile-replace**. The compile-replace algorithm was developed to account for non-concatenative morphological processes, such as the vocalisation patterns in Arabic and full reduplication in Malay (Beesley and Karttunen, 2003). The compile-replace algorithm for reduplication works by delimiting the portion of the network that is affected by compile-replace. This so-called ‘portion’ of the net is defined as a regular expression and is delimited by the tags ‘^[]’ and ‘^[]’ on the lower side of the net and ‘Redup[]’ and ‘]Redup[]’ on the upper side. When the compile-replace algorithm is invoked, the net defined by regular expression between ‘^[]’ and ‘^[]’ is copied. There are computational limitations to what can be defined within these delimited tags, so in practice we apply compile replace to predefined lexemes, or stems, as listed in the LEXC stem lexicon, with optional predefined affixes, and exclude unknown stems.

3.3 Word Level Parser: XLE

The tool used for parsing, XLE, only utilises two of the three levels of representations discussed earlier: f-structure and c-structure. In Figure 1 both the ‘Word Parser’ and ‘Sentence Parser’ utilise XLE.

XLE is a grammar development environment which interprets grammars written in an electronic parseable variation of LFG. It is the tool used for defining the phrase structure, as well as the sublexical rules, which describes how the word is com-

posed. We construct these rules via c-structure rules, which look like traditional grammar rewrite rules but with annotations giving us the information that can only be encoded via the phrase structure. Within the “Word Parser” component, there are defined **sublexical rules** that are interpreted using XLE.

This component crucially relies on the analysis of the ‘Morphological Analyser’ and its output must be a meaningful representation of the input, which is the surface form of the reduplicated verb. There is a semantic motivation for wanting to represent the predicates in (1) *menarik-menarik*, (2) *menarik-narik*, and (3) *tarik-menarik* in different ways. We would want our morphological analysis to be sensitive to their semantic differences, however small or large. For these given predicates, there are three important components of the word to represent:

- reduplication: Redup[]Redup
- the agentive voice affix: AV
- the verb stem: *tarik* “pull”

We could represent the analysis of *menarik-menarik* as Redup[AV+tarik]Redup, but we would want to differentiate *menarik-narik* from this and so could represent this as AV+Redup[tarik]Redup. However, this also seems a plausible output for *tarik-menarik*, as does the former. In order to enforce a unique representation for all three, we arrive at:

(16) *menarik-menarik*: Redup[AV+tarik]Redup

(17) *menarik-narik*: AV+Redup[tarik]Redup

(18) *tarik-menarik*: Redup[tarikAV+]Redup

The first reduplicated example, *menarik-menarik* in (16), with the stem *tarik* “pull” means “pull again and again”. The second example, *menarik-narik*, has a very similar meaning to (16), but the major difference is that the action (i.e. the “pulling” in the case of *tarik*) is repeated faster. The last example *tarik-menarik*, (18), means “pull at each other”, in a tug-of-war fashion.

4 Integration into the Grammar

4.1 Reciprocals

From a formal point of view, it seems that the reciprocal is formed by marking two verbs with un-

dergoer and agentive voice, which forms a linking between the agent and the patient of the action. In Indonesian, undergoer voice is the unmarked bare verb as shown by Arka and Manning (2008), and agentive voice is marked with *meN*. This compound verb analysis gives us an adequate semantic account of reciprocals, but more needs to be done in order to explain the arity reduction of the resulting predicate, as seen in (19) where *mereka* “they” is the only argument of the verb.

(19) *Mereka pukul-memukul*
 they UV.hit-AV+hit
 “They hit each other”

We adopt a similar analysis of reciprocals in Indonesian to the analysis of Alsina (1997) and Butt (1997) for causative verbs in Chichewa and permissives in Urdu, respectively: the reciprocal verb formation in Indonesian is a type of complex predicate in that the elements of the reciprocal combine to alter the argument structure of the resulting predicate, which acts as a single grammatical unit (Alsina et al., 1997). Even though the same principle of predicate composition applies, these analyses do not involve valence reduction as it does in Indonesian, but rather valence increasing.

Although the undergoer plus agentive voice treatment of reciprocal formation gives us a neat account of argument linking, these verb stems would then be considered two separate verbs as they both have their own voice marking, and therefore have their own values for the VOICE attribute in their f-structure attribute value matrices. This means, from an implementation point of view, there would have to be a semantic identity check to ensure both verbs have the same verb stem. For this implementation reason, we choose to keep this as a process within the ‘Morphological Analyser’ and as reduplication rather than verb compounding. This then saves a form of ‘identity matching’ of the two stems at a later stage.

The reciprocal is interpreted as such by virtue of the reduplication construction where the agentive voice affix *meN* is inserted between the reduplicated stems. Therefore the ‘instructions’, if you will, for composing reciprocals are encoded in the sublexical c-structure rules and manifested in the f-structure, as it affects argument linking.

If we step back from the implementation for a moment, we can represent schematically what happens to the arguments of a regular transitive verb such as (20), when it is composed as a reciprocal (21). But what we want is to create a general rule that allows this operation to apply to all transitive verbs where the resulting reduplicated form has an interpretable reciprocal predicate.

(20) *pukul* < agent, patient >

(21) *pukul-memukul* < agent&patient >

The important components of the reciprocal word forming sublexical rules are as follows:

- The input to the rule has one argument (ARG), which is a transitive stem verb that requires a subject (ARG SUBJ) and an object (ARG OBJ)
- The resulting complex predicate (RECIP-rocal) only requires a subject (SUBJ) that must be plural (NUM pl)

The input predicate ARG must still be complete, meaning that it must still satisfy its (ARG SUBJ) and (ARG OBJ), which is the **agent** and **patient** in (20). That is, the verb on which the RECIPROCAL verb is formed is transitive and requires all its arguments to be filled. We can achieve this via coindexing the subject and object of the input predicate ARG with the subject of the derived predicate RECIP.

(22) RECIP < (SUBJ_i), ARG < (SUBJ_i), (OBJ_i) > >

The resulting predicate is mono-valent, in that it only needs to satisfy a subject, however it has an input predicate. Figure 4 shows the resulting f-structure for the reciprocal sentence in (19). The first line (labelled PRED) is the representation of the semantics of the head of the attribute value matrix over which it has scope. In this case the PRED on the first line represents the main verb *pukul-memukul* “hit-reciprocally”. It tells us it is a derived reciprocal whose first slot is satisfied by the attribute value matrix labelled 4, which is the subject; the second slot is satisfied by a verb that takes two arguments.

The c-structure for (19) is shown in Figure 5. Each of the numbered nodes corresponds to a component in the f-structure. It is clear in the c-structure

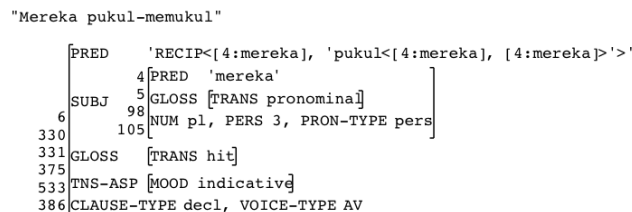


Figure 4: Feature structure

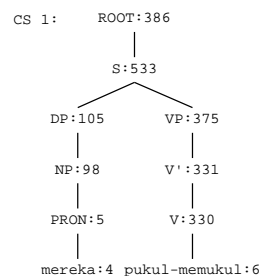


Figure 5: Constituent structure

that the verb only takes one noun phrase argument, which is the subject. The operation that composes the derived reciprocal verb requires a transitive verb as input, which is *pukul* “hit” in Figure 4, and it is represented in the f-structure inside the PRED value for the RECIP verb.

4.2 Distributed Reduplication

The implementation of the non-reciprocal reduplication is less involved, in that this construction simply triggers an additional feature in the f-structure, however it has its complexities too. The main issue is: what feature should be added?

We discussed earlier that reduplication constructions such as (23) are not exactly the same as the English progressive aspect, and in some examples have more of an iterative aspect, in that the action is repeated but not necessarily with one sustained action over time, but in a start-stop fashion. Therefore a feature such as ITER + as part of the tense-aspect definition of the clause could be added to the f-structure.

Noun phrases in Indonesian are underspecified for number, much like the English noun phrases that are headed with mass nouns, such as *rice*. However the

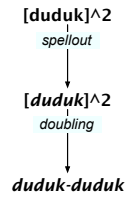


Figure 6: Spellout then doubling of *duduk*

reduplication on the verb can impose a plural reading on the argument(s) of the verb, where the action is applied to each and every member of the argument of the verb, as seen in the second translation in (23) ((12) is an earlier example).

- (23) *Dia memukul-mukul temannya*
 He AV+hit-hit his.friend
 “He was (repeatedly) hitting his friend.”/“He hit each of his friends.”

When the verb determines the number of its arguments, this is called a **pluractional** verb (Corbett, 2000). Pluractionality specifies that the action is over multiple affected objects, and so we could add the attribute-value pair PLURACT + for these constructions, which would not be part of the tense-aspect definition of the clause.

In the present implementation, for sentences such as (23), both solutions are possible.

5 Rejigging the Morphological Analyser

Traditional analyses of reduplication have been modelled on a theory of phonological copying or a doubling of a phonologically-rendered form. This entails that we begin with a lexeme *duduk* “sit”, we then execute the spellout rule or the phonological rendering giving us *duduk* /dudu?/, and then this form is doubled producing *duduk-duduk* “sit around”, as seen in Figure 6.

The architecture of the Pisceldo et al. (2008) morphological analyser in Figure 3 models this idea of how the reduplication mechanism works. Specifically, the morphophonemic rules are executed first, giving us our spelled-out rendering, which is then doubled. Certainly when we examine some of the morphophonemic facts of reduplication in Indonesian, it gives support for this architecture. Such an



Figure 7: Examples where spellout must precede doubling



Figure 8: Examples where the order of double and spellout has no consequence

example is shown in Figure 7, which is the realisation of AV+[tarik]^2 (agentive voice prefix with the reduplicated stem *tarik* “pull”); Figure 8 presents a case where relative ordering does not matter.

However, this implementation cannot account for the full morphophonemic facts of reduplication, namely the reciprocal construction, without the aid of corrective spellout rules.

We see in Figure 9 that for these types of examples we need to allow for the doubling of the verbs stem, ensuring appropriate attachment of voice marking to the respective stems, before we allow for spellout to take place. The notation (-,AV) is an indication of how the voice affixes are ‘multiplied out’ upon reduplication.

Inkelas and Zoll (2005) puts forward a theory of reduplication, Morphological Doubling Theory (MDT), that can incorporate both strategies allowing spellout and doubling in any order, and that both strategies are called for. They also claim that the reduplicated stems are a lot more discrete and can bear different affixes, and their phonological rendering can be realised independently from each other. This seems to model what we observe in the reciprocal construction in Indonesian: an independence of phonological realisation. The two different ordering for spellout and doubling very neatly separates

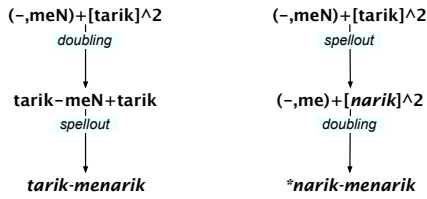


Figure 9: Examples where Doubling must precede spell-out

out the two types of reduplication processes. Therefore within both the morphological analyser and the sublexical component, reciprocal reduplication and distributive reduplication are handled aptly as distinct separate processes, as seen in Figure 10.

Although we do not in whole borrow from MDT, some of the concepts put forward in the theory gave us cause to see the two reciprocal processes as being separate in the morphological analyser. As such, we have allowed for both spellout before reduplication and then spelling out this doubling process. We see these two processes as serving different purposes: one for the aspectual/distributed reduplication and the other for the reciprocal reduplication. It seems apt to be treating them differently in the morphological analyser, given that they are implemented so differently in the sublexical word building component.

6 Conclusion

In this study, we discussed reduplication in combination with the voice marker AV. There are other voice prefixes such as the passives *di*, *ter* and *ber* that we still need to investigate. We would want to see whether these would require special treatment. In addition, we need to investigate more deeply the interaction with applicative morphology such as *-kan* and *-i*, as shown in (24), and to ensure that we develop an analysis that would complement our existing implementation of the applicatives (Arka et al., 2009).

- (24) *Mereka beli-membilkan mobil*
 they AV+beli-beli+KAN car
 “They bought cars for each other”

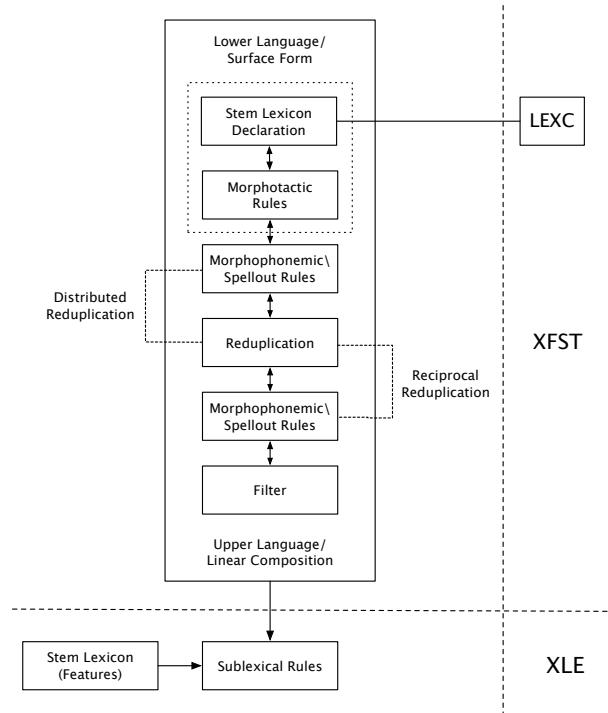


Figure 10: Current morphological analyser with separated doubling process for the two types reduplication constructions.

We had initially considered all reduplication in the morphological analyser as the same doubling process, and implemented reduplication accordingly. Although the two forms of reduplication we were investigating, reciprocal and distributive, were morphosyntactically very different and so had to be implemented very differently in the sublexical component, we had not considered handling them differently from each other in the morphological analyser to account for their differences with respect to their morphophonemic facts. Instead of preemptive corrective rules, we implemented another component to correctly treat the stems of the reciprocal reduplication and distributive reduplication as being more independent of each other, with respect to their phonological realisation.

References

- Alex Alsina, Joan Bresnan, and Peter Sells. 1997. Complex predicates: Structure and theory. In Alex Alsina,

- Joan Bresnan, and Peter Sells, editors, *Complex Predicates*, pages 1–12. CSLI, Stanford, USA.
- Alex Alsina. 1997. Causatives in Bantu and Romance. In Alex Alsina, Joan Bresnan, and Peter Sells, editors, *Complex Predicates*, pages 1–12. CSLI, Stanford, USA.
- I Wayan Arka and Christopher Manning. 2008. Voice and grammatical relations in Indonesian: a new perspective. In Simon Musgrave and Peter Austin, editors, *Voice and grammatical relations in Austronesian languages*, pages 45–69. CSLI, Stanford, USA.
- I Wayan Arka, Avery Andrews, Mary Dalrymple, Meladel Mistica, and Jane Simpson. 2009. A computational morphosyntactic analysis for the applicative -i in Indonesian. In *Proceedings of LFG2009*.
- Kenneth R. Beesley and Lauri Karttunen. 2003. *Finite State Morphology*. CSLI Publications.
- Joan Bresnan. 2001. *Lexical Functional Syntax*. Blackwell, Massachusetts, USA.
- Miriam Butt, Tracy Holloway King, Maria-Eugenia Nino, and Frederique Segond. 1999. *A Grammar Writers Cookbook*. CSLI, Stanford, USA.
- Miriam Butt. 1997. Complex predicates in Urdu. In Alex Alsina, Joan Bresnan, and Peter Sells, editors, *Complex Predicates*, pages 1–12. CSLI, Stanford, USA.
- Greville Corbett. 2000. *Number*. Cambridge University Press, Cambridge, UK.
- Sharon Inkelas and Cheryl Zoll. 2005. *Reduplication: Doubling in Morphology*. Cambridge Studies in Linguistics, 106. Cambridge University Press, Dunno, USA.
- Simon Musgrave. 2008. Introduction: Voice and grammatical relations in austronesian languages. In Simon Musgrave and Peter Austin, editors, *Voice and grammatical relations in Austronesian languages*, pages 1–21. CSLI, Stanford, USA.
- Femphy Pisceldo, Rahmad Mahendra, Ruli Manurun, and I Wayan Arka. 2008. A two-level morphological analyser for Indonesian. In *Proceedings of the Australasian Language Technology Association Workshop*, volume 6, pages 88–96.
- James Neil Sneddon. 1996. *Indonesian reference grammar*. Allen Unwin, St. Leonards, N.S.W.