



Prihantoro  
[prihantoro@lancaster.ac.uk](mailto:prihantoro@lancaster.ac.uk)

# The Annotation Of Indonesian Reduplications With Nooj



*ESRC Centre  
Corpus Approach to Social Science  
Department of Linguistics  
Lancaster University  
United Kingdom*

# Morpheme level annotation for Indonesian with Nooj

Preferences

General Lexical Analysis Syntactic

Lexical Resources for: id

Priority	Resource
H9	EN-v2-dyka.nod
H8	EN-v2-yumi.nom
H7	EN-v2-yumi-LV2.nom
H6	EN-v2-dito_LV1.nom
H5	EN-v2-dito_LV2.nom
H4	EN-v2-dito_LV3.nom
H3	EN-v2-dito_LV4.nom
H2	EN-v2-dito_LV5.nom
H1	EN-v2-dito_LV6 .nom
L1	EN-v2-dyka-foreign.nod

General Lexical Analysis Syntactic Analysis

Syntactic Resources:

Order	Grammar
1	EN-v2-pri.r.nog
2	EN-v2-pri_LV1.nog
3	EN-v2-pri_LV2.nog
4	EN-v2-pri_LV3.nog
5	EN-v2-pri_LV4.nog

# Morpheme level annotation for Indonesian with Nooj

The screenshot shows the Nooj interface with the following components:

- Show Text Annotation Structure**: A checkbox at the top left.
- Raw word form**: The main text area containing "Dalam jumpa pers bersama Dirut Bursa Efek Jakarta BEJ, Erry Firmansyah di gedung Depkeu Jakarta, Rabu malam, Menkeu menje beberapa faktor ekonomi Indonesia justru membaik. Kita melihat faktor inflasi dari makanan akan segera hilang yang terkonfirmasi dari harga beras di semua propinsi, katanya. Demikian pula dengan harga minyak."
- Implementation code**: A menu on the right with options:
  - Linguistic Analysis
  - Locate Pattern... (Ctrl+L)
  - Export colored text as an RTF document
  - Export annotated text as an XML document...** (selected)
  - Modify Text (Ctrl+M)
- Annotations**: Below the raw text, the word "justru" is annotated with ID 162 and category X+Dyka. The prefix "mem" is annotated with ID 169 and category PFX+meN+R:VER+Active+ID=112303. The suffix "baik" is annotated with ID 169.01 and category baik,ADJ+ROOT+L4+aK+TX+Dyka.

Orthographic form  
(allomorph)  
*mem*

Formal categorial analysis  
*PFX*

Citation form  
(morpheme)  
*meN*

Functional analysis  
*R:VER+Active*

Tweak annotations > Nooj XML Export >  
modify to this by using small PHP script)

```
290
291 <word form="justru">
292 ~justru @justru +ADV+ROOT
293 </word>
294
295 <word form="membaik">
296 ~mem @meN +PFX+R_VER+Active
297 ~baik @baik +ADJ+ROOT
298 </word>
299
300 <word form=".">>
301 ~. @. +PUNC+D+Ldic
302 </word>
303 }
```

# Indonesian reduplication: type and productivity

## Full reduplication

*pukul-pukul*

hit-RED.Full.ltrv

'to hit iteratively'

## Partial reduplication

*te-tua*

RED.Part.Nomzr-old

'elder'

## Imitative reduplication

*warna-warni*

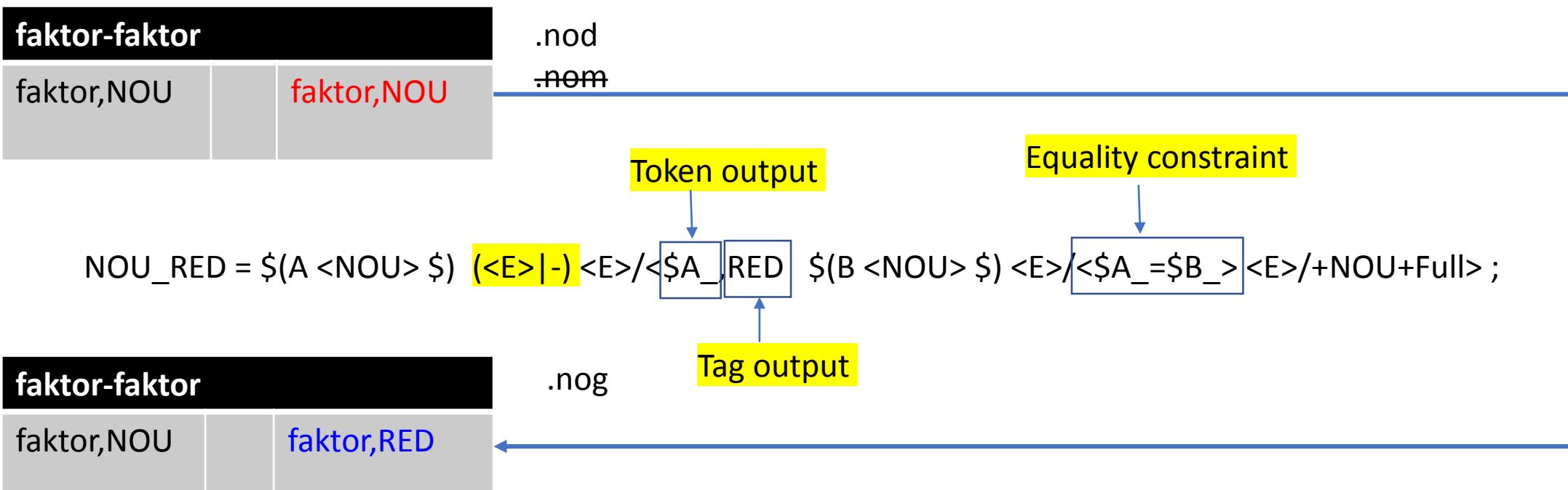
Color-RED.Adjvzr.Var

'colorful'

pukul,VER+ROOT+L5+aK+T1+Dyka  
pula,ADV+ROOT+L4+aK+TX+Dyka  
pulang,VER+ROOT+L6+aK+T0+Dyk;  
pulas,ADJ+ROOT+L5+aK+TX+Dyka  
pulau,NOU+ROOT+L5+aK+TX+Dyka  
pulen,ADJ+ROOT+L5+aK+TX+Dyka  
pulih,VER+ROOT+L5+aK+T0+Dyka  
pulpen,NOU+ROOT+L6+aK+TX+Dyk  
pulsa,NOU+ROOT+L5+aK+TX+Dyka  
puluh,NUM+ROOT+L5+aK+TX+Dyka  
pulung,NOU+ROOT+L6+aK+TX+Dyk  
puma,NOU+ROOT+L4+aK+TX+Dyka

terang=benderang,<terang,ADJ+ROOT+Dyka><benderang,RED+Imitative+ADJ+Dyka>+UNAMB  
tetamu,<te,RED+Partial+Dyka><tamu,NOU+ROOT+Dyka>+UNAMB  
tetangga,<te,RED+Partial+Dyka><tangga,NOU+ROOT+Dyka>+UNAMB  
tetapi,<te,RED+Partial+Dyka><tapi,KNJ+ROOT+Dyka>+UNAMB  
tetikus,<te,RED+Partial+Dyka><tikus,NOU+ROOT+Dyka>+UNAMB  
tindak=tanduk,<tindak,NOU+ROOT+Dyka><tanduk,RED+Imitative+NOU+Dyka>+UNAMB  
utak=utik,<utak,RED+Imitative+Dyka><utik,VER+ROOT+Dyka>+UNAMB  
dedaunan,<de,RED+Partial+Dyka><daun,NOU+ROOT+Dyka><an,SFX+Dyka>+UNAMB  
terang=benderang,<terang,ADJ+ROOT+Dyka><benderang,RED+Imitative+ADJ+Dyka>+UNAMB  
jejaring,<je,RED+Partial+Dyka><jaring,NOU+ROOT+Dyka>+UNAMB  
pontang=panting,<pontang,RED+Imitative+Dyka><panting,VER+ROOT+Dyka>+UNAMB  
kekasih,<ke,RED+Partial+Dyka><kasih,NOU+ROOT+Dyka>+UNAMB  
tetua,<te,RED+Partial+Dyka><tua,ADJ+ROOT+Dyka>+UNAMB  
leluhur,<le,RED+Partial+Dyka><luhur,ADJ+ROOT+Dyka>+UNAMB  
lelaki,<le,RED+Partial+Dyka><laki,NOU+ROOT+Dyka>+UNAMB  
tetamu,<te,RED+Partial+Dyka><tamu,NOU+ROOT+Dyka>+UNAMB  
iejaka,<ie,RED+Partial+Dyka><jaka,NOU+ROOT+Dyka>+UNAMB  
warna=warni,<warna,NOU+ROOT+Dyka><warni,RED+Imitative+NOU+Dyka>+UNAMB

# The annotation of reduplication



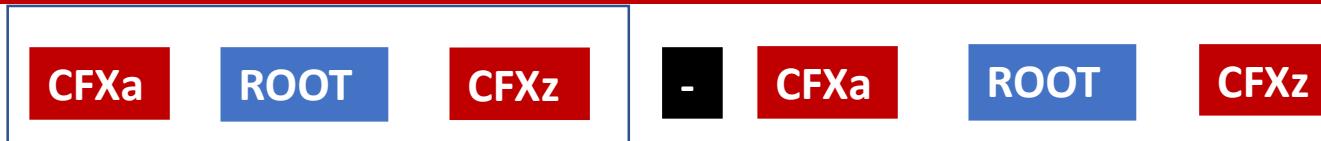
Rabu 10/1 hanya efek sementara dari faktor-faktor regional seperti penurunan harga komoditi perkebunan dan Dalam jumpa pers bersama Dirut Bursa Efek Jakarta BEJ, Erry Firmansyah di gedung Depkeu Jakarta, Rabu m Kita melihat faktor inflasi dari makanan akan segera hilang yang terkonfirmasi dari harga beras di semua proj

+L4+aK+TX	34 faktor,NOU+ROOT+L6+aK+TX	41 faktor,RED+NOU+Full	48 regional,ADJ+ROOT+L8+aK+TX
-----------	--------------------------------	---------------------------	----------------------------------

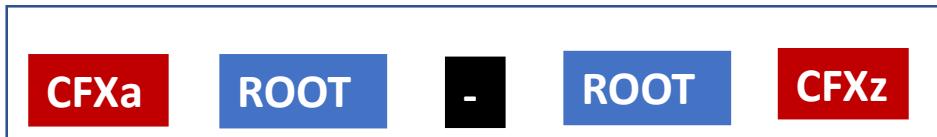
# Circumfix + reduplication type



Type 1



Type 2



Selain kata pukul-memukul untuk menyatakan makna 'saling memukul' juga digunakan kata berpukul-pukulan.

(2) Setelah berpukul-pukulan, tendangan samping Johan mengakhiri pertahanan Rudi.

Karena pukul-memukul dan berpukul-pukulan sudah mengandung makna peristiwa berbalasan, kata saling tidak perlu ditambahkan di depan kata.

(3) Setelah saling pukul-memukul beberapa saat, mereka mengalami kecapaian.

(4) Setelah saling berpukul-pukulan beberapa saat, mereka mengalami kecapaian.

Jika kata saling di depan kata itu dipertahankan, makna 'saling' yang terkandung dalam kata kerja harus dihilangkan dengan cara mengubah

12 12.01

ber,PFX+beR+R:VER+R11.2.1 pukul,VER+ROOT+L5+aK+T1

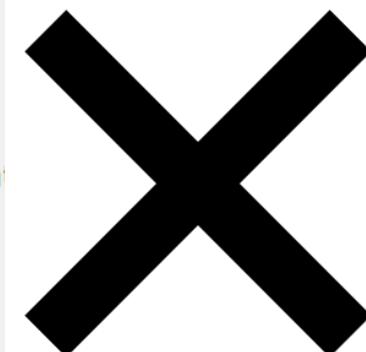
21

pukul,RED+ROOT+VER+Full+Reciprocal+Iterative+FUNGSI an,SFX+R:NOU+R6.5

21.01

CFXa

CFXz



# Possible solutions

	Possible solutions	Type	Complexities for users	Internal modification of Nooj engine
1	Unification of .nom and .nog	Automatic	Low	Required
2	Modification of .nom to accept hyphen	Automatic	Low	Required
3	Add rules in current .nom: process the circumfix like prefix + suffix	Automatic	High	Not necessary
4	Encode – as an entry in .dic so that will be ‘read’ as letter in .nom	Automatic	Low	Not necessary
5	Hard encoding of morphological analysis in .dic	Manual	Low	Not necessary

My choice now

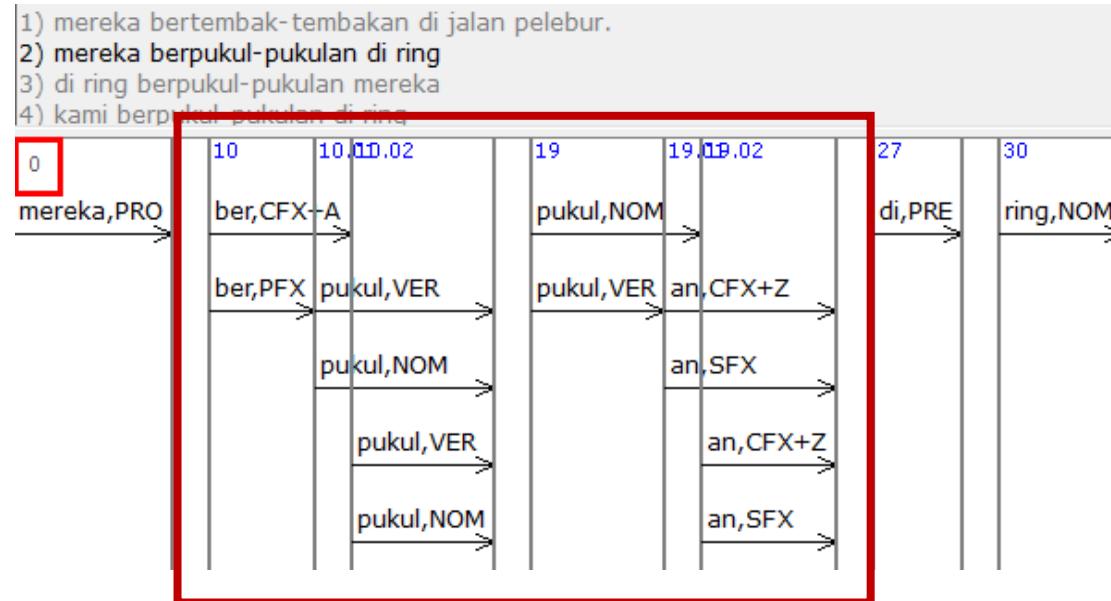
Does not work

Have not experimented at large

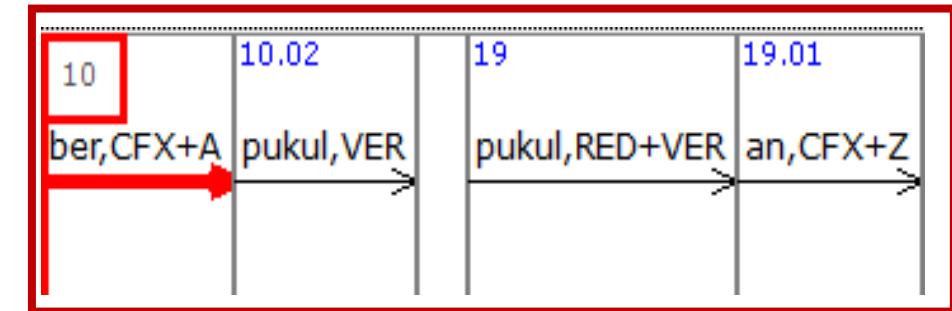
I simply cant do

# Option 3: main target

Current state



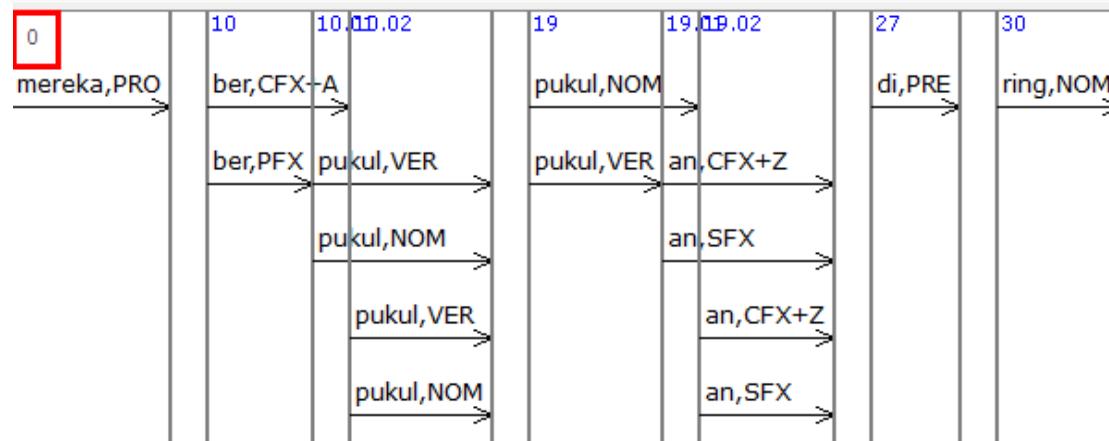
Expectation



# Step 1: - <ROOT> root into - <RED>

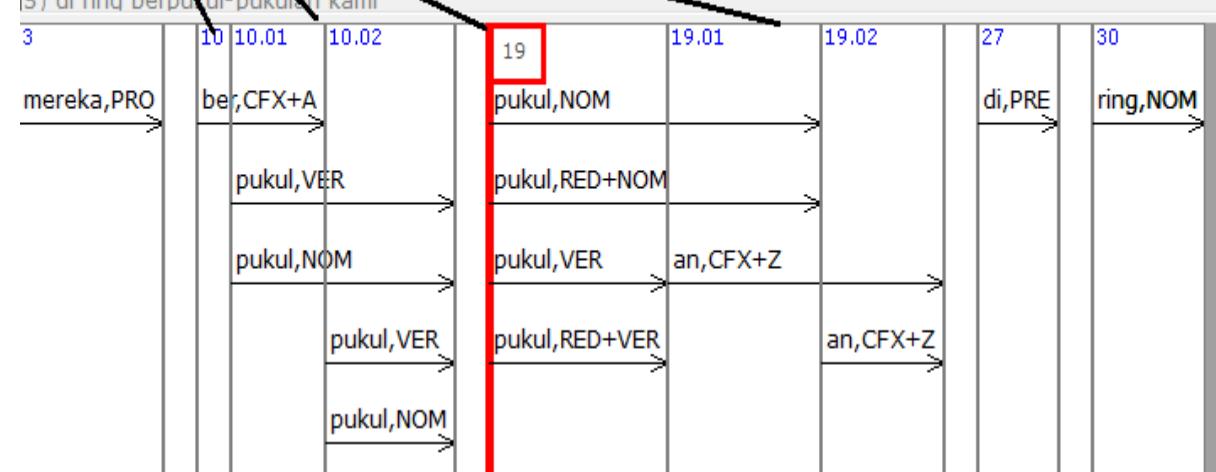
Current state

- 1) mereka bertembak-tembak di jalan pelebur.
- 2) mereka berpukul-pukulan di ring
- 3) di ring berpukul-pukulan mereka
- 4) kami berpukul-pukulan di ring



Next state

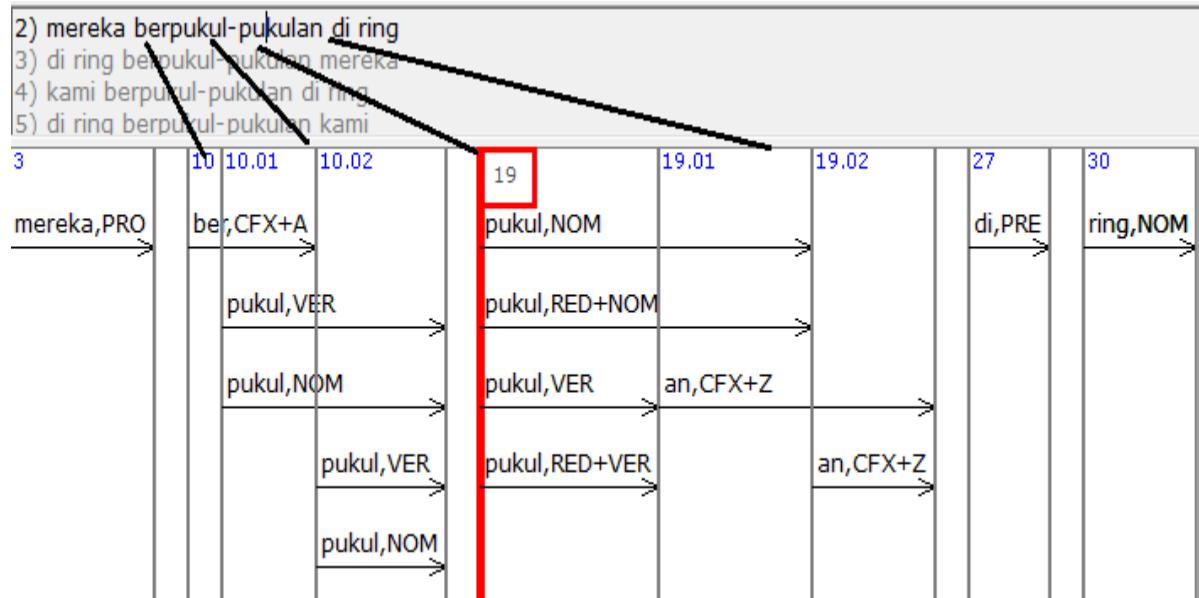
- 2) mereka berpukul-pukulan di ring
- 3) di ring berpukul-pukulan mereka
- 4) kami berpukul-pukulan di ring
- 5) di ring berpukul-pukulan kami



<CFX>/<CFX> \$(A <NOM> \$) - <E>/<\$A\_,RED+NOM \$ (B <NOM> \$) <E>/<\$A\_=\$B\_> <E>/> <CFX>/<CFX>;  
 <CFX>/<CFX> \$(A <VER> \$) - <E>/<\$A\_,RED+VER \$ (B <VER> \$) <E>/<\$A\_=\$B\_> <E>/> <CFX>/<CFX>;

# Step 2: disambiguate into <CFX><VER>-<RED><CFX>

Current state



Next state



<CFX> \$(A <VER>/<VER> \$) - \$(B <RED+VER>/<RED+VER> \$) <E>/<\$A\_=\$B\_> <CFX>

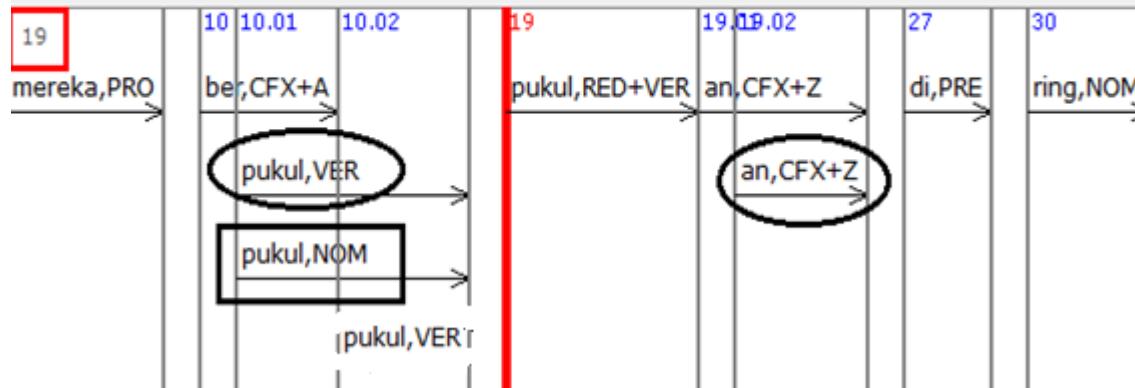
Wrong

orphan

# Step 3: Remove wrong annotations

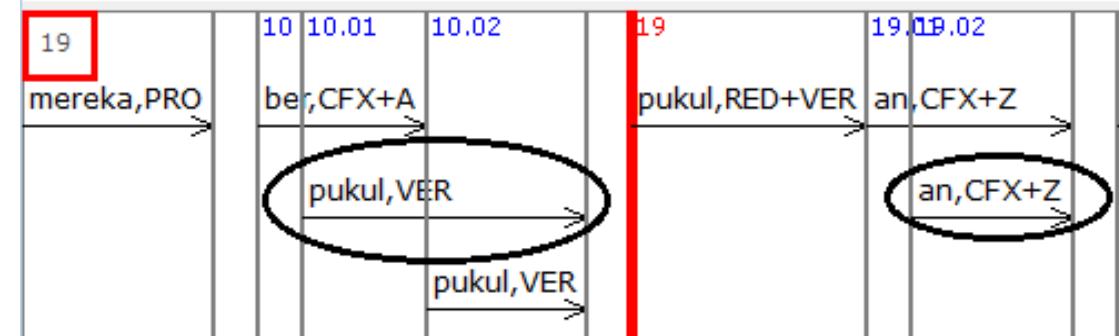
Current state

- 1) mereka bertembak-tembakan di jalan pelebur.
- 2) mereka berpukul-pukulan di ring
- 3) di ring berpukul-pukulan mereka
- 4) kami berpukul-pukulan di ring



Next state

- 1) mereka bertembak-tembakan di jalan pelebur.
- 2) mereka berpukul-pukulan di ring
- 3) di ring berpukul-pukulan mereka
- 4) kami berpukul-pukulan di ring

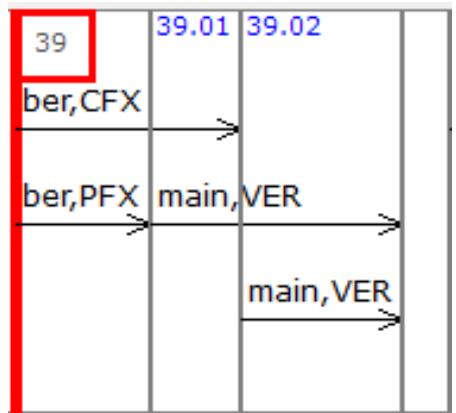


$\$(A \text{ <NOM>/<VER>} \$) - \$(B \text{ <RED+VER>/<RED+VER>} \$) \text{ <E>/<\$A\_=\$B\_>}$

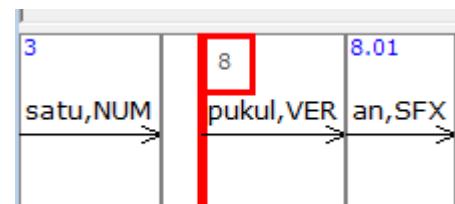
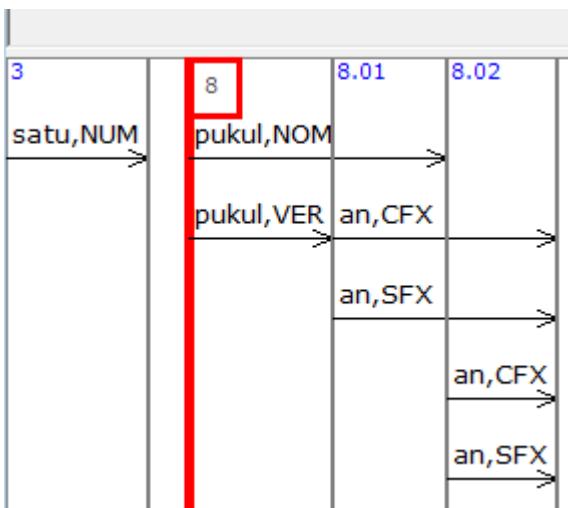
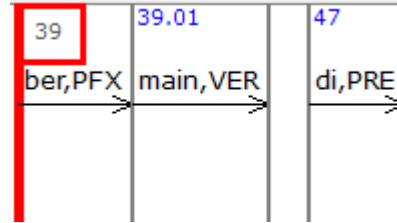
< target correct annotation>/<output correct annotation> does not work!  
So, < target wrong annotation>/<output correct annotation>

# Entailment problem

Current state

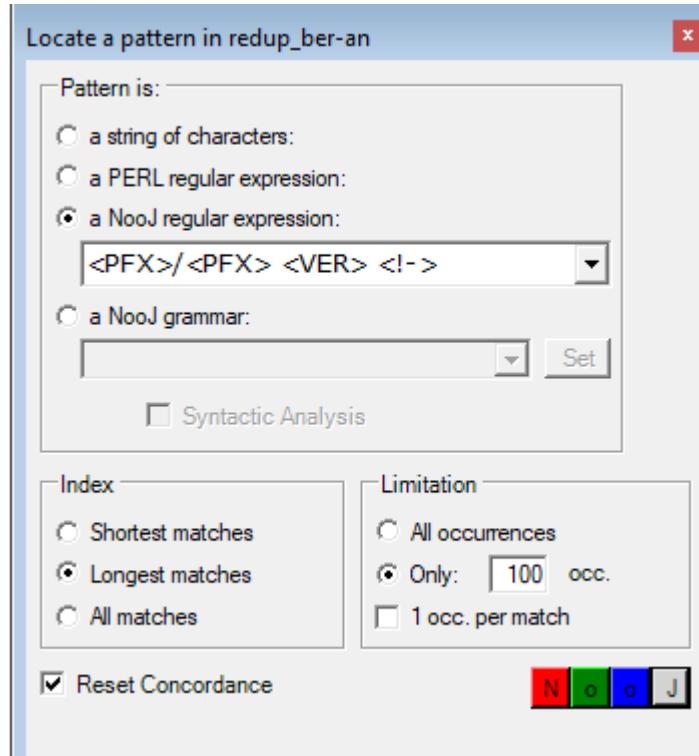
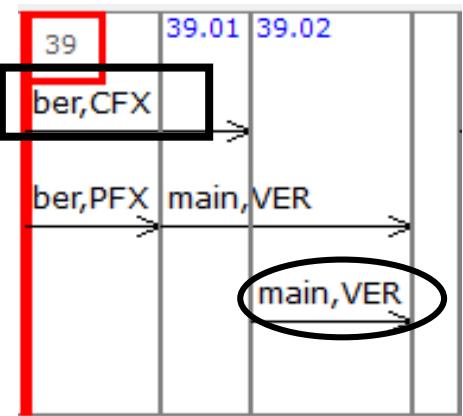


Expectation

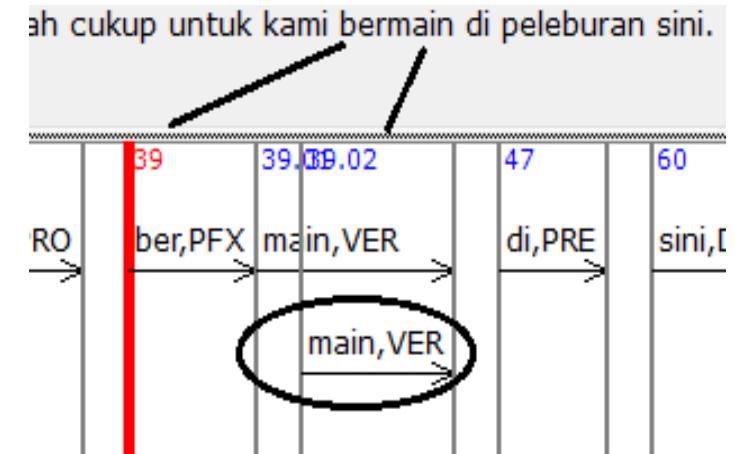


# Step 4: disambiguate into <PFX><VER>

Current state

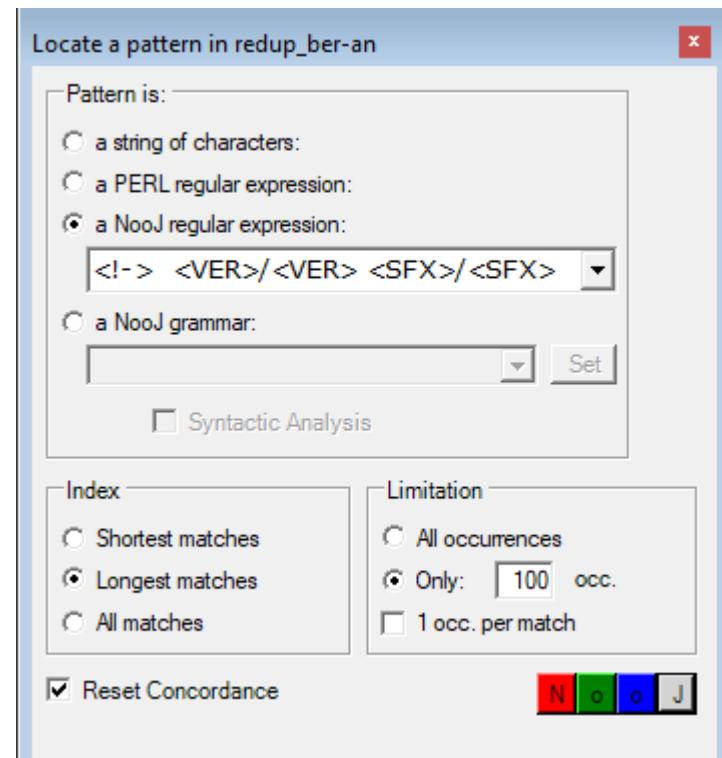
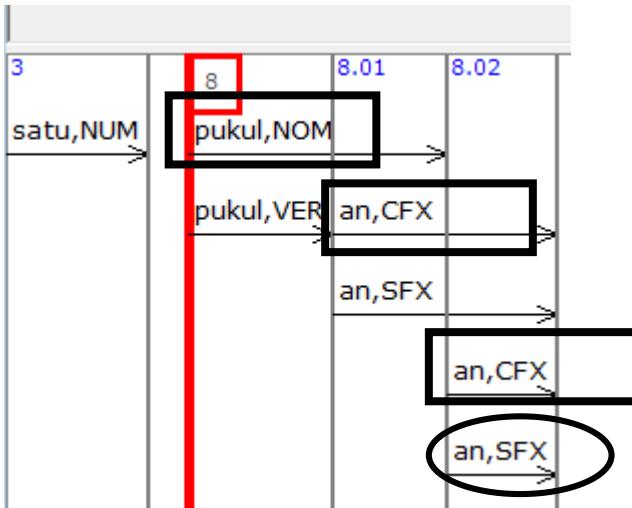


Next state

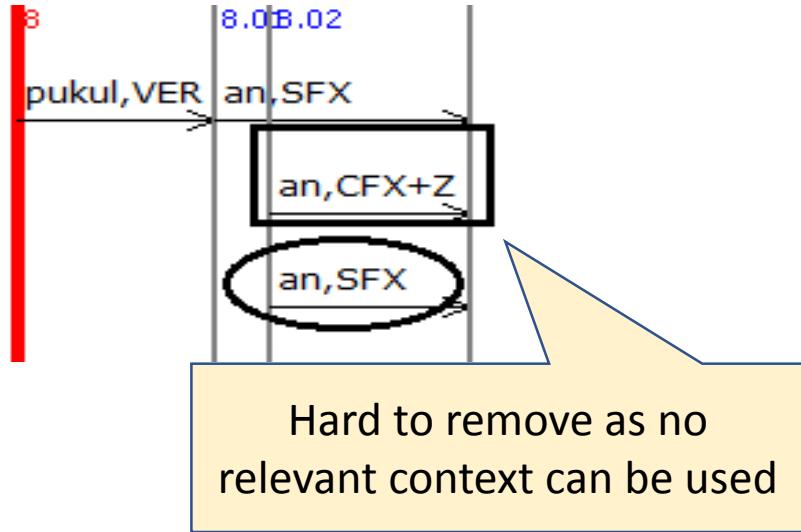


# Step 5: disambiguate into <VER><SFX>

Current state

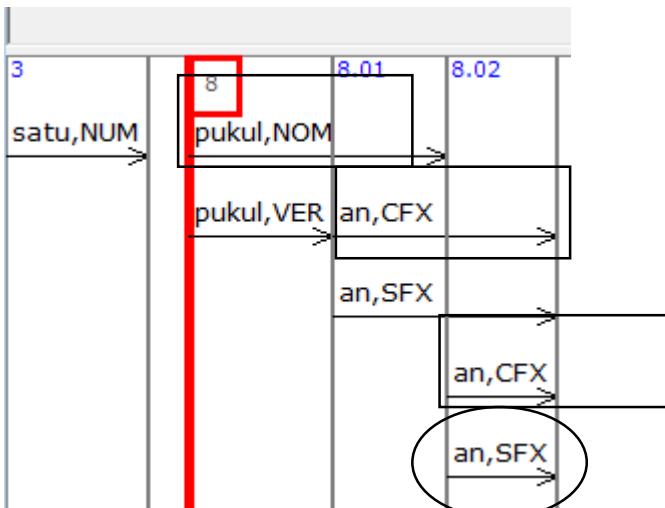


Next state



# Step 6: revise step 5

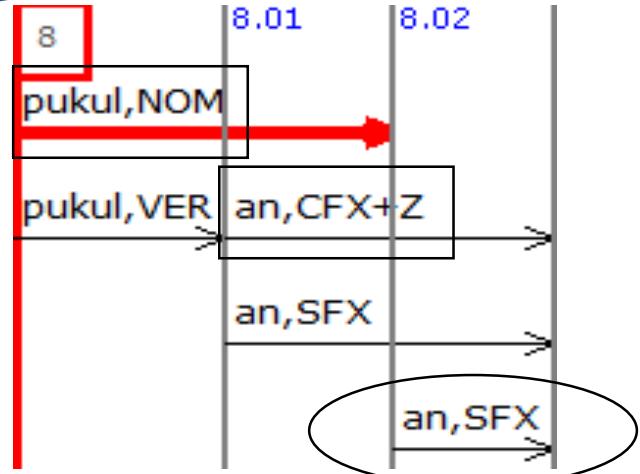
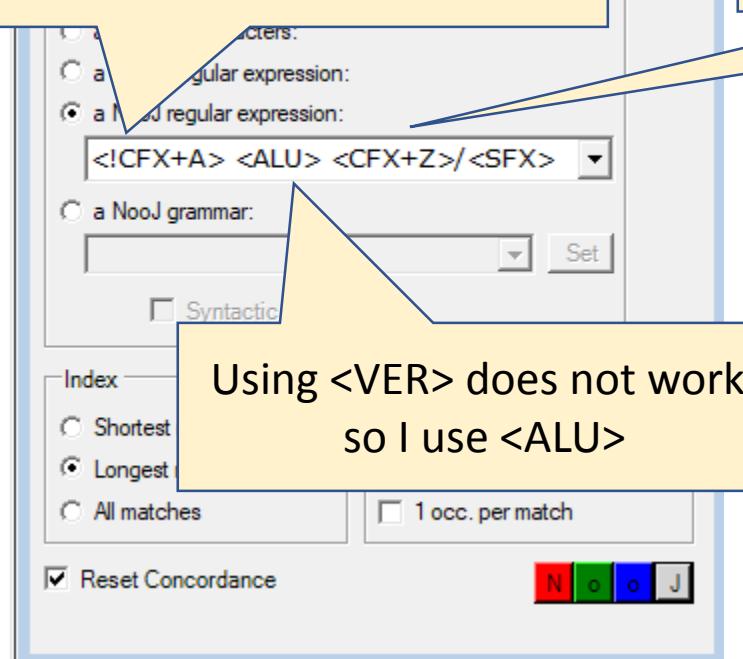
Current state



Using <!-> does not work; so,  
I use <!CFX+A>

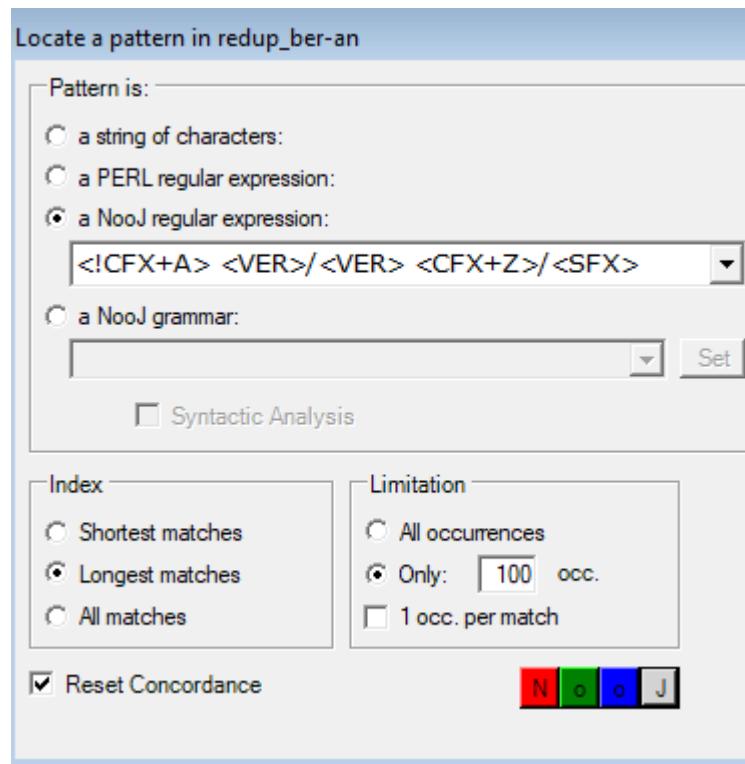
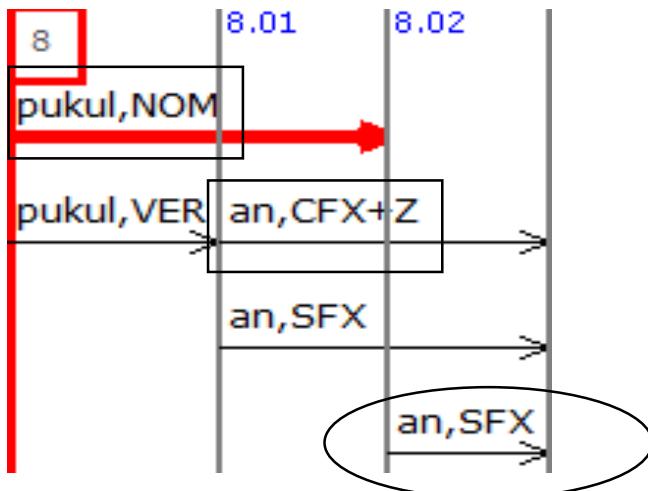
Using <VER> does not work;  
so I use <ALU>

< target wrong annotation>/<output correct  
annotation>

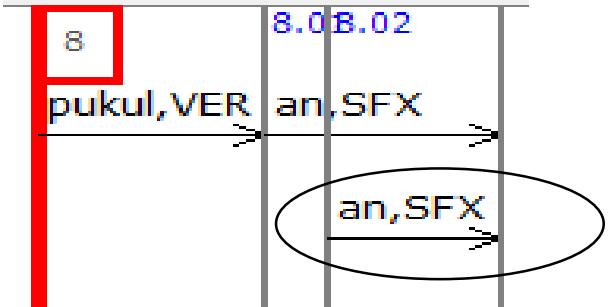


# Step 7: remove wrong annotations

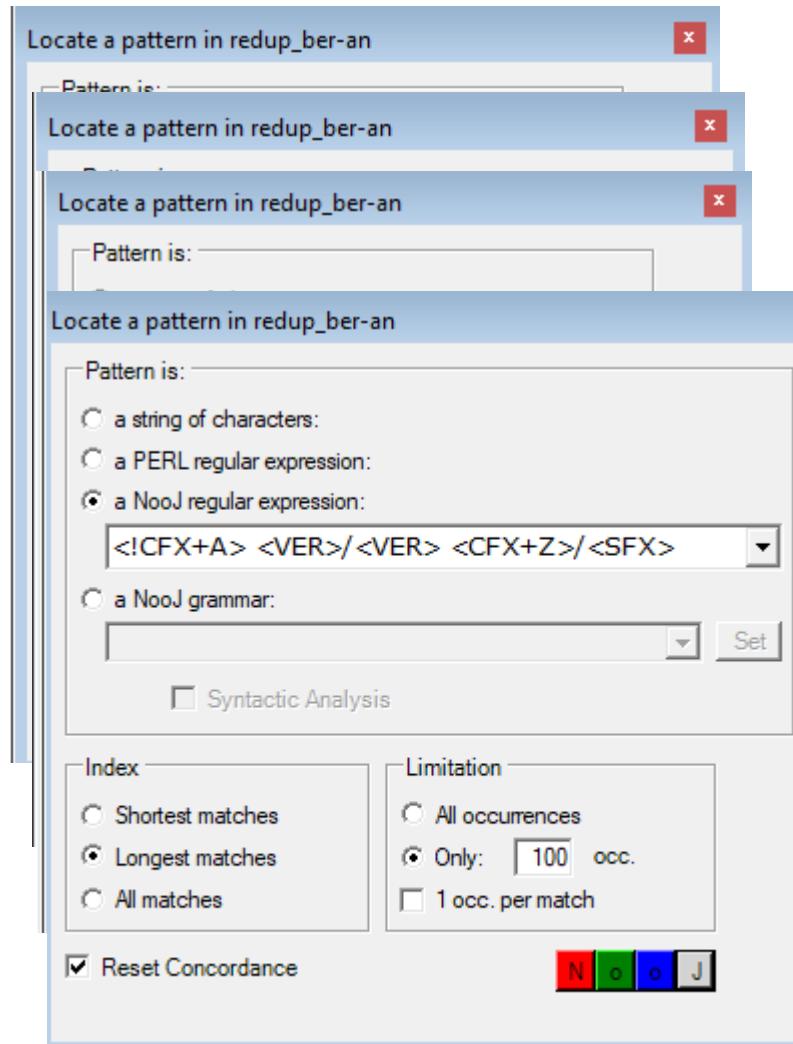
Current state



Next state



# Manual sequence VS Preference setting



VS

The screenshot shows the "Preferences" window with the "Syntactic Analysis" tab selected. Under "Syntactic Resources", there is a table titled "Lexical Resources for: id". A large blue arrow points from the left towards this table.

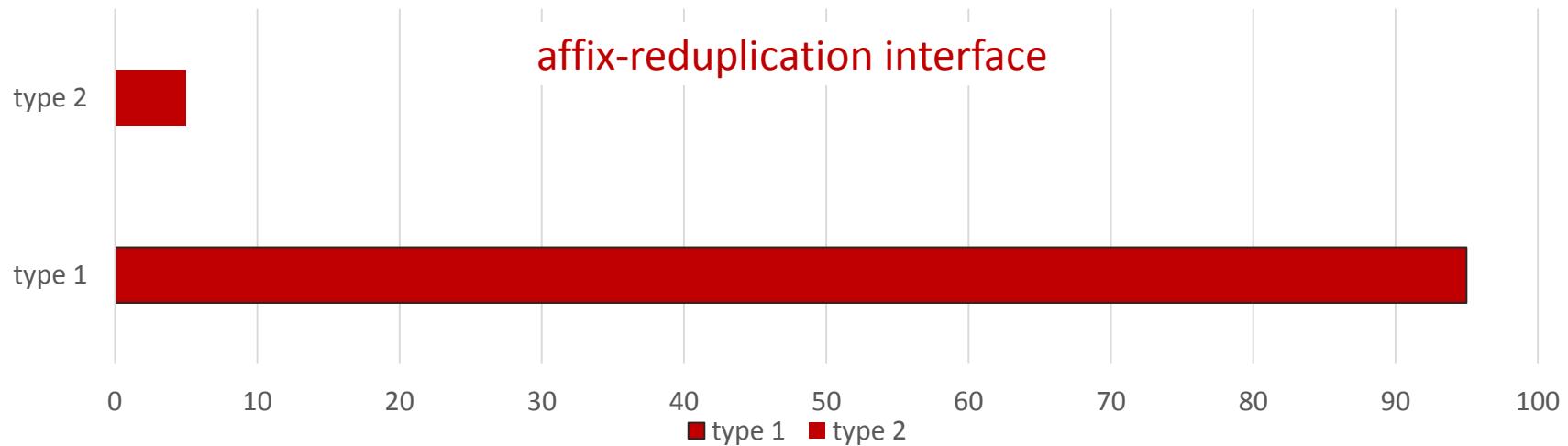
Order	Grammar
1	EN-v2-pri.r.nog
2	EN-v2-pri_LV1.nog
3	EN-v2-pri_LV2.nog
4	EN-v2-pri_LV3.nog
5	EN-v2-pri_LV4.nog

Priority	Resource
H9	EN-v2-dyka.nod
H8	EN-v2-yumi.nom
H7	EN-v2-yumi-LV2.nom
H6	EN-v2-dito_LV1.nom
H5	EN-v2-dito_LV2.nom
H4	EN-v2-dito_LV3.nom
H3	EN-v2-dito_LV4.nom
H2	EN-v2-dito_LV5.nom
H1	EN-v2-dito_LV6.nom
L1	EN-v2-dyka-foreign.nod

Does not work

# Results

- Relatively successful annotations of reduplications
- Problems with discontinuous reduplication-affixation interface
  - Problematic on ambiguous roots
- Effective and simple treatment requires modification of Nooj engine
- Complexity of current experiment (option 3) is high



- Other challenging aspect: Encode reduplication function (plural number, variation, iterative aspect, reciprocal voice, etc)



Prihantoro

[prihantoro@lancaster.ac.uk](mailto:prihantoro@lancaster.ac.uk)

A word cloud collage where the words 'Thank You' are repeated in various languages. Other words like 'Merci', 'Danke sehr', 'Arigato', 'Xie Xie', 'Aho', 'Toda', 'Mesi', etc., are also present.

Questions?