Serial Verb Constructions in Indonesian: An HPSG Analysis and Its Computational Implementation

David Moeljadi and Viola Ow

Division of Linguistics and Multilingual Studies, Nanyang Technological University, Singapore

Chulalongkorn International Student Symposium on Southeast Asian Linguistics, Maha Chakri Sirindhorn Building, Faculty of Arts, Chulalongkorn University

9 June 2017
1. Serial Verb Constructions in Indonesian

2. Indonesian data

3. Analysis and computational implementation

4. Conclusion and future work
Serial Verb Construction (SVC)

- a type of syntactic feature
- a sequence of **two or more juxtaposed verbs**
- **a single monoclausal** structure
- **neither** of which is an **auxiliary**
- referring to a **single** (possibly **complex**) **event**
- **not** separated by any **conjunctions**
- share **at least one semantic argument**
- has **a single intonation contour**
- encompass **a broad range of semantic relationships**, based on verb meaning and context
- **no overt** syntactic **marking**, conjunction, or other morpheme to **indicate the semantic relation** between the verbs
- **different languages** impose **different restrictions** as to which specific **combinations of verbs** are permissible

(Kroeger 2004 [9], Aikhenvald and Dixon 2006 [2])
SVC in Indonesian

- at least two adjacent verbs
- without intervening material (occurring *contiguously*, focusing on the surface form of the constructions)
- refer to closely related events
- occur in the same, **single intonation unit**: ‘the intonation properties of a single verb clause, and not a sequence of clauses’ (Aikhenvald 1999: 470)\[1\]
- **share at least one argument**

Englebretson (2003: 128-133)[7]

Indonesian reference grammars (Sneddon et al. 2010[16], Alwi et al. 2014[3], Mintz 2002[11]) note this phenomenon but do not mention clearly that it is SVC.
Indonesian language

- bahasa Indonesia “the language of Indonesia”
- Austronesian language family > Western Malayo-Polynesian language > Malayic > Standard Malay, Indonesian, other Malay varieties [10]
- the sole official and national language of the Republic of Indonesia, the common language for hundreds of ethnic groups in Indonesia [3]
- L1 speakers: around 43 million
  - L2 speakers: more than 156 million (2010 census data)
- Latin script
- Morphologically mildly agglutinative: prefixes, suffixes, ...
- SVO word order, nominative-accusative alignment pattern
- Diglossic: “High” and “Low” varieties [13] [15]
- We analyzed SVC in the Standard “High” variety of Indonesian (cf. Englebretson (2003) uses a corpus of colloquial “Low” variety)
Serial verbs as putative complements

1 **control** construction (Arka 2000): a relation of referential dependence between an unexpressed argument in an embedded clause (controlled argument) and an expressed or unexpressed argument (the controller) in a matrix clause.

V1 = the head, ‘control verbs’, complement-taking predicates/modality verbs (Englebretson 2003)/commitment verbs and orientation verbs (Arka 2000); V2 = (in)transitive

**commitment verbs**: mencoba “try”, menolak “refuse”, berusaha “attempt”, mulai “begin”, ...

**orientation verbs**: ingin “desire”, berhak “to have rights”, perlu “need”, suka “like (to do something)”, tahu “know how to”, ...

(1) *Budi mencoba mengejar Adi.*
*Budi meN-try meN-chase Adi*
“Budi tries to chase Adi.”
SVC in Englebretson (2003) and other previous works II

raising (Arka 2000): the argument that is ‘thematically’ associated with an embedded clause is syntactically expressed as the argument of the matrix verb, where the matrix verb does not assign any thematic role to the ‘raised’ argument

V1 = the head, ‘raising verbs’; V2 = (in)transitive

(2) *Budi tampak mengejar Adi.*
Budi appear meN-chase Adi
“It appears that Budi chases Adi.”
Serial verbs with other semantic relationships (Englebretson 2003)

1 **Manner** serialization
   V1 = head, intransitive; V2 = (in)transitive
   V2 expresses how V1 is done

(3) *Budi berjalan menggunakan tongkat.*
   Budi walk mEN-use stick
   “Budi walks using a stick.”

2 **Purpose** serialization
   V1 = head, intransitive; V2 = (in)transitive
   V1 enables V2 to happen, V1 must be done first

(4) *Budi pulang mengambil uang.*
   Budi go.home mEN-take money
   “Budi goes home to get money.”
3. Periphrastic **causative** construction
usually non-standard in Indonesian and does not occur in formal Indonesian, except for the lexicalized *beri tahu* “inform” (*beri* “give”, *tahu* “know”)

(5) *Budi kasih mati mereka.*
Budi give die 3PL
“Budi kills them.”

4. **Coordinated actions**
V1 and V2 occur rapidly and repetitively, appearing to be simultaneous
V1 = (in)transitive, V2 = (in)transitive

(6) *Mereka memukul mengeroyok Budi.*
3PL meN-beat meN-gang.up Budi
“They beat up and gang up Budi.”
Other semantic relationships not mentioned in Englebretson (2003)

1. ‘origin’ or ‘source’ (Alwi et al. 2014)[3]
   - V1 = head, intransitive; V2 = (in)transitive
   - V1 happens after V2

(7)  \textit{Budi pulang bertamasya.}
    Budi go.home picnic
    “Budi goes home from picnic.”
    (ambiguous, can mean “Budi goes home to picnic”)

2. ‘resultative’ (Kridalaksana 1989)[8]
   - a combination of an action (predicated by V1, the head) and a result (predicated by V2) which is caused by that action
   - V1 = transitive and V2 = intransitive

(8)  \textit{Budi membunuh mati Adi.}
    Budi meN-kill die Adi
    “Budi kills Adi until Adi dies.”
Extracting SVCs from corpus

- The Nanyang Technological University — Multilingual Corpus (NTU-MC) [17]
  - A parallel English-Chinese-Japanese-Indonesian corpus
  - 2,975 Indonesian sentences
  - Three sources: Singapore tourism board website, Sherlock Holmes short story, Japanese short story
  - Part-of-speech (POS) tagged → V-V can be extracted

- The PARSEME annotation guidelines: state-of-the-art for annotation of verbal multi-word expressions [5]
  - We modified the guidelines for English and wrote suggested PARSEME guidelines for Indonesian SVCs.
  - For each extracted V-V:
    - If V1 is a control verb like mencoba “try”, V1+V2 is a control SVC
    - If V1 is a raising verb like tampak “appear”, V1+V2 is a raising SVC
    - If V2 expresses the manner of V1, V1+V2 is a manner SVC
    - If V2 indicates the purpose of V1, V1+V2 is a purpose SVC, etc.
## Distribution of Indonesian SVCs in corpus

<table>
<thead>
<tr>
<th>Type of SVC</th>
<th>Number</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>control</td>
<td>21</td>
<td>72.0%</td>
</tr>
<tr>
<td>manner</td>
<td>3</td>
<td>10.3%</td>
</tr>
<tr>
<td>raising</td>
<td>2</td>
<td>6.9%</td>
</tr>
<tr>
<td>purpose</td>
<td>2</td>
<td>6.9%</td>
</tr>
<tr>
<td>coordinated action</td>
<td>1</td>
<td>3.4%</td>
</tr>
<tr>
<td>source</td>
<td>0</td>
<td>0.0%</td>
</tr>
<tr>
<td>resultative</td>
<td>0</td>
<td>0.0%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>29</strong></td>
<td><strong>100.0%</strong></td>
</tr>
</tbody>
</table>

**Table:** Distribution of Indonesian SVCs in the corpus

(out of 45 possible Indonesian SVCs extracted)
Examples of Indonesian SVCs found in corpus I

(9)  \[\ldots Holmes \underline{mencoba} \underline{membuka} \underline{palang} \underline{itu} \ldots\]
Holmes meN-try meN-open shutter that
“\ldots Holmes \ldots tries to open that shutter \ldots” (SID: 10417) control

(10)  \[\ldots saya \underline{pulang} \underline{melalui} \underline{halaman} \underline{itu} \ldots\]
1sg go.home pass.through yard that
“\ldots I go home by passing through that...yard.” (SID: 10500) manner

(11)  \[\underline{Waktu} \underline{terasa} \underline{berlalu} \underline{dengan} \underline{lambat} \underline{sekali}.\]
time feel pass with slow very
“It feels that time passes very slowly.” (SID: 10585) raising
(12)  

...tangannya menggapai-gapai mencari pertolongan...
hand-3sg reach.out seek help
“...his hands reach out to seek help...” (SID: 10183) **purpose**
(ambiguous, can be classified into **coordinated action**)  

(13)  

Saya segera berlari menuju kamar ayah tiri kami...
1sg soon run head.towards room father step- 1pl.excl
“I soon run towards our stepfather’s room...” (SID: 10193) **coordinated action**
HPSG and MRS

- **Head Driven Phrase Structure Grammar (HPSG):** unification- and constraint-based grammar [14]
- **Minimal Recursion Semantics (MRS):** semantic framework in which the semantic structures are underspecified for scope [6]
- **Deep Linguistic Processing with HPSG Initiative (DELPH-IN):** a research collaboration between linguists and computer scientists which builds and develops open-source grammars etc. using HPSG and MRS (http://www.delphin.net)
- **Indonesian Resource Grammar (INDRA):** open-source Indonesian computational grammar within DELPH-IN [12]
  
  http://moin.delph-in.net/IndraTop
Analyses

Indonesian SVCs

- serial verbs as putative complements
- Other semantic relationships
- control (V1 = control verbs)
- raising (V1 = raising verbs)
- V1 = intransitive (manner, purpose, source, coordinated actions)
- V1, V2 = transitive, shared object (coordinated actions)
- V1 = transitive, V2 = intransitive (resultative)
Control and Raising SVCs

- **Control SVC**
  - Control verbs assign a semantic role to their subject
  - *Holmes mencoba membuka palang itu*  
    “Holmes tries to open that shutter”

- **Raising SVC**
  - Raising verbs do not assign a semantic role to their subject
  - *Waktu terasa berlalu*  
    “It feels that time passes”
SVCs with other semantic relationships

- V1=intransitive, V2=(in)transitive (manner, purpose, source, coordinated actions)
  - *Saya pulang melalui halaman itu*
    - “I go home by passing through that yard”

![Diagram of SVC with semantic relationships]
• V1, V2 = transitive with a shared object (coordinated actions)
  ▶ *Tangannya menggapai-gapai mencari pertolongan*
  “His hands reach out and seek help”
SVCs with other semantic relationships III

- V1=transitive, V2=intransitive (resultative meaning)
  - *Budi membunuh mati Adi*
    - “Budi kills Adi until Adi dies”
Conclusion and future work

- Our analyses of Indonesian SVCs depart from Englebretson (2003) with references from Arka (2000) and other reference grammars.
- We wrote a Python script to extract SVCs from NTU-MC and annotated the types of SVCs with our modified PARSEME guidelines.
- We analyzed and modeled Indonesian SVCs using HPSG and MRS and implemented our analyses in INDRA.
- For control and raising verbs, we posit lexical types.
- For verbs with other semantic relationships, we made rules based on the transitivity of V1 and V2 and the shared arguments.
- There may be a closed class of verbs that occurs as V1 or V2 in each semantic relationship → we plan to get more corpus data and do more work on verb subcategorization.
Acknowledgements

- Thanks to Sanghoun Song for his help in implementing the control and raising SVCs in the early stage
- Thanks to Francis Bond, Frantisek Kratochvil, and Hiroki Nomoto for the precious comments and advice


