

SE463

Tutorial 2:
State Machine Diagram for Assigning
Positions (or Shapes) to Arabic,
Persian, or Urdu Characters

Ali Niknafs

September 20th, 2011

Inputs

Letters

Arabic letters have four positions/shapes:

- Standalone
- Connect After
- Connect Previous
- Connect Both

Letters

CP	CB	CA	SA
ف	ف	ف	ف
ك	ك	ك	ك

Letters

Some letters have only two shapes.

For these:

CP = CB and CA = SA

Example:

CP	CB	CA	SA
L	L	l	l

Inputs

Delimiters

Space tab , . : ; ? ! etc

Digits

. 1 2 3 4 5 6 7 8 9

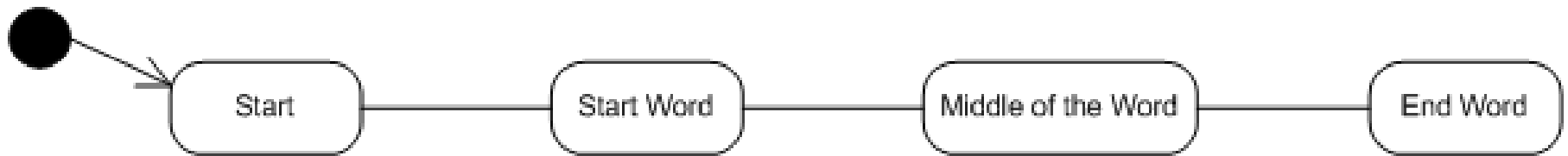
Inputs

Zero width non-joiner (ZWNJ)

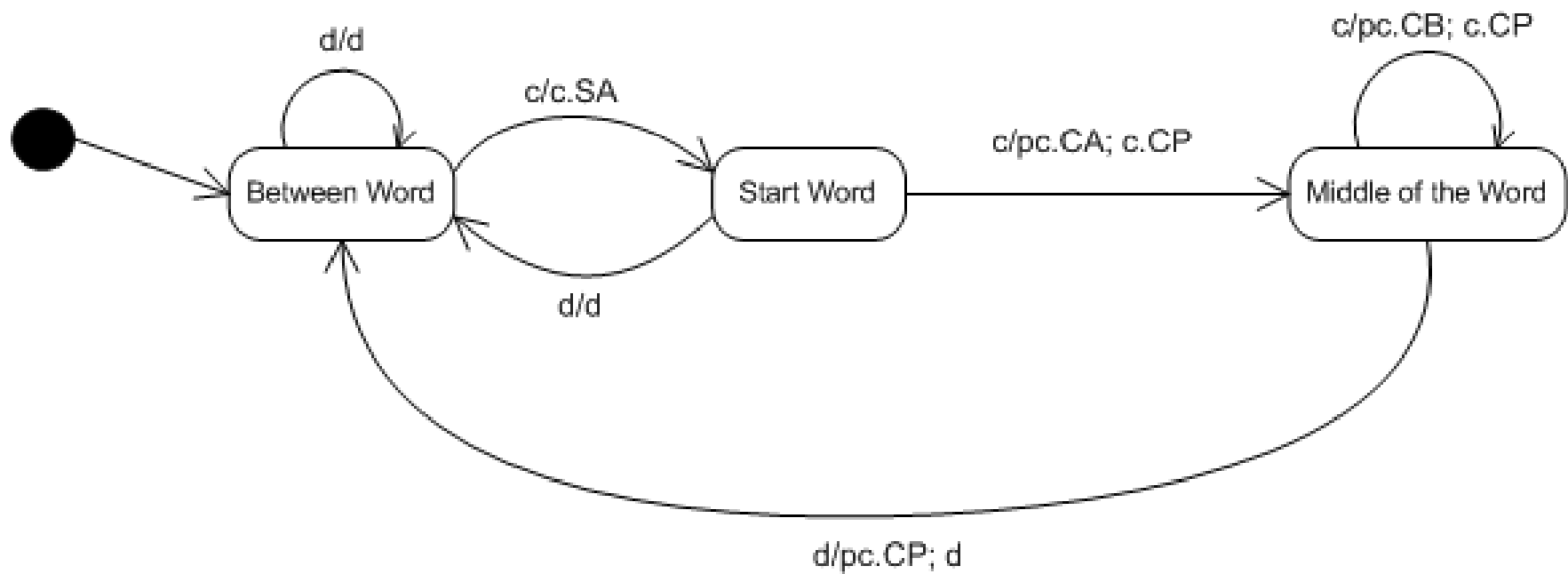
Without ZWNJ: بِل

With ZWNJ: بِل

What are the states?

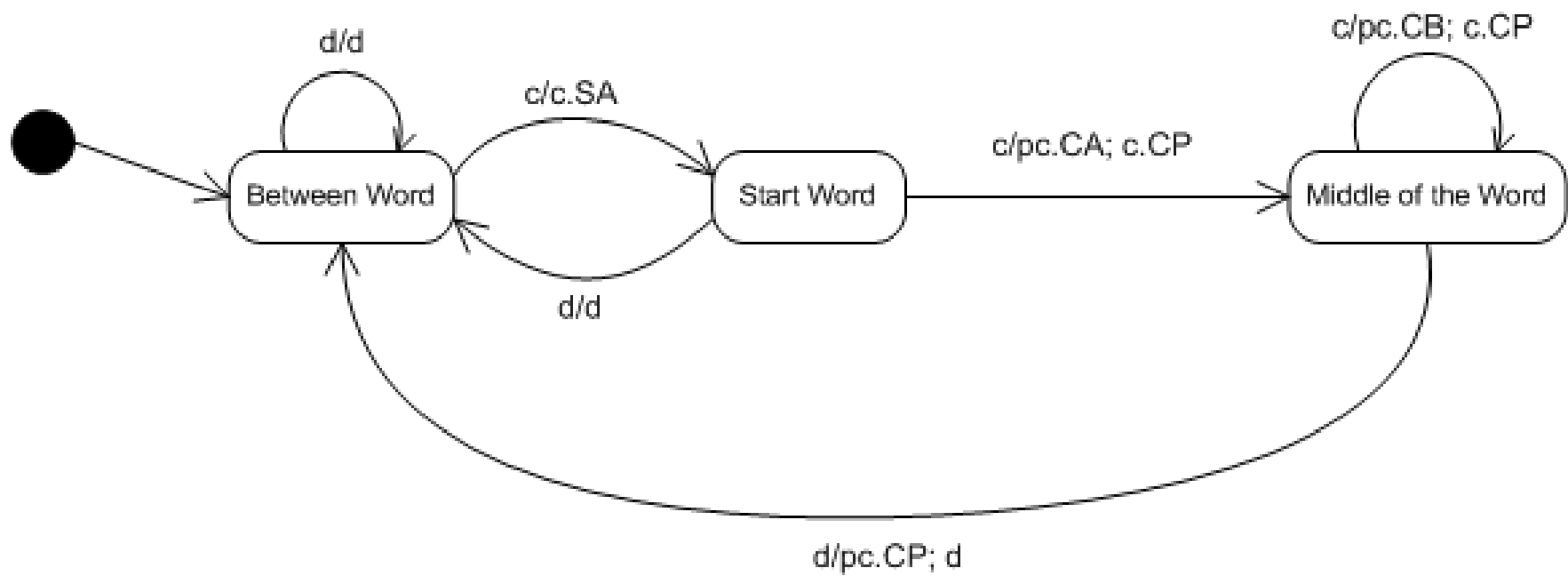


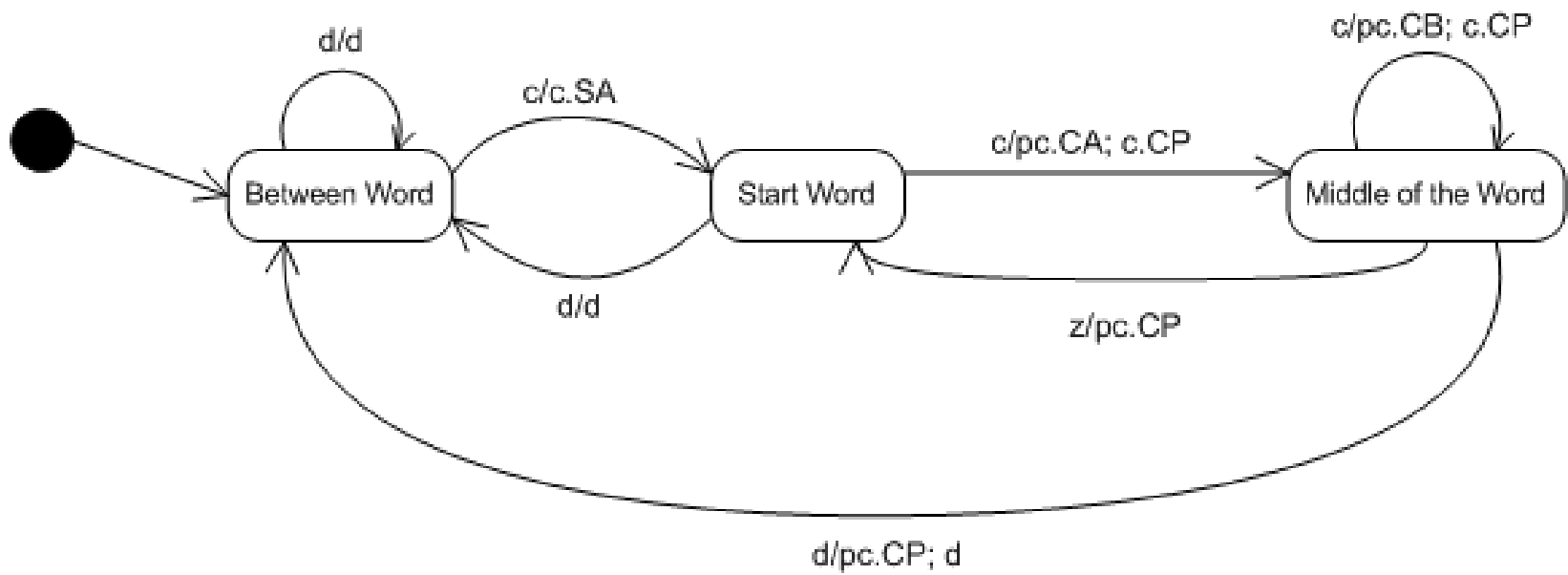
- d: delimiter
- c: character



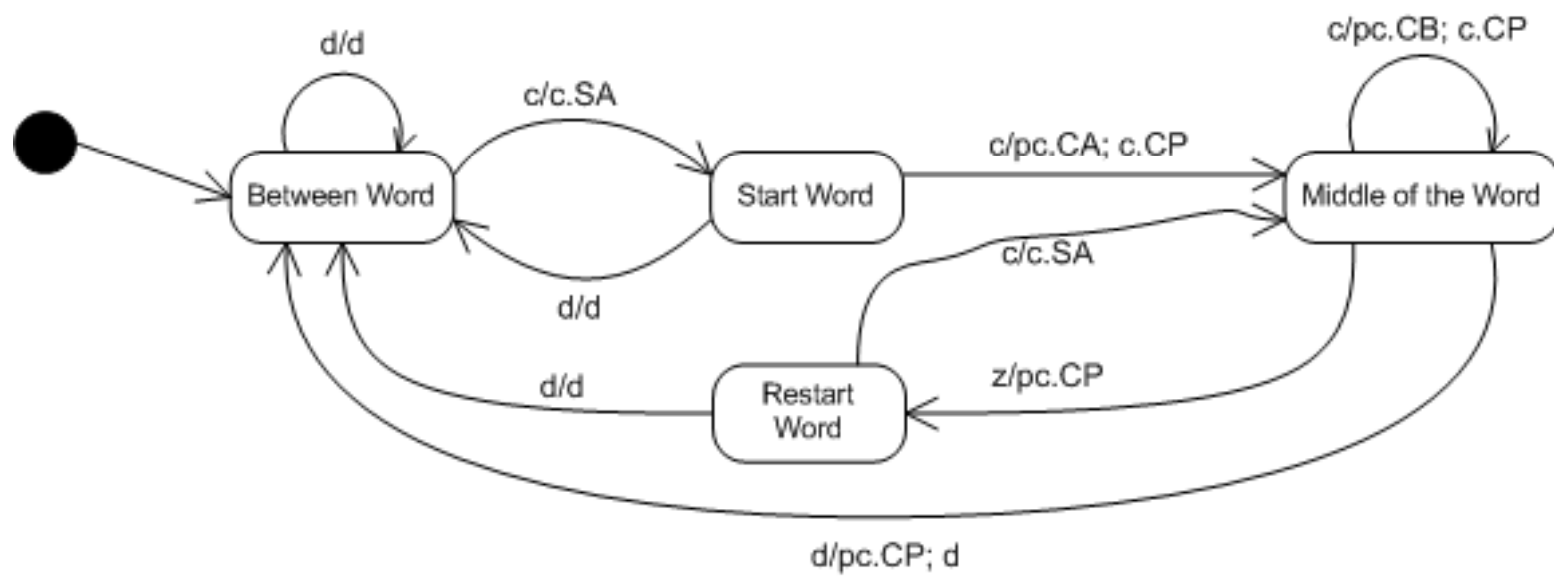
Sample: `_bbb_`

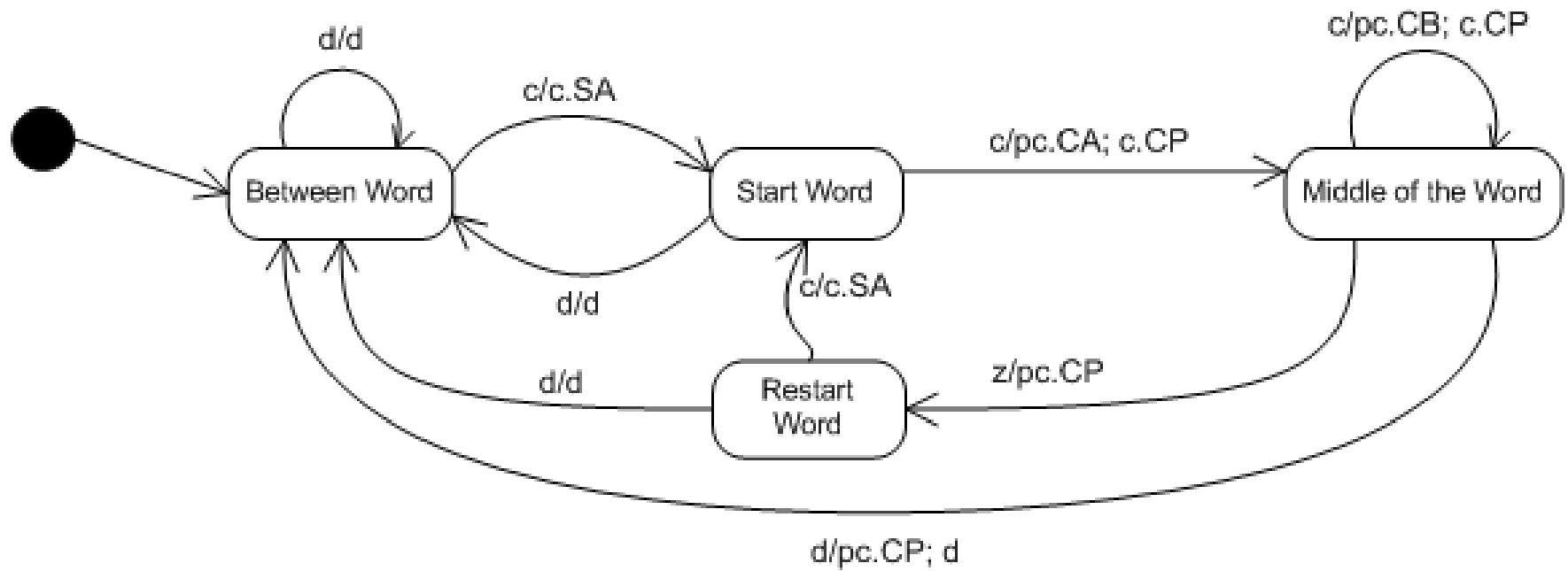
- What about ZWNJ?



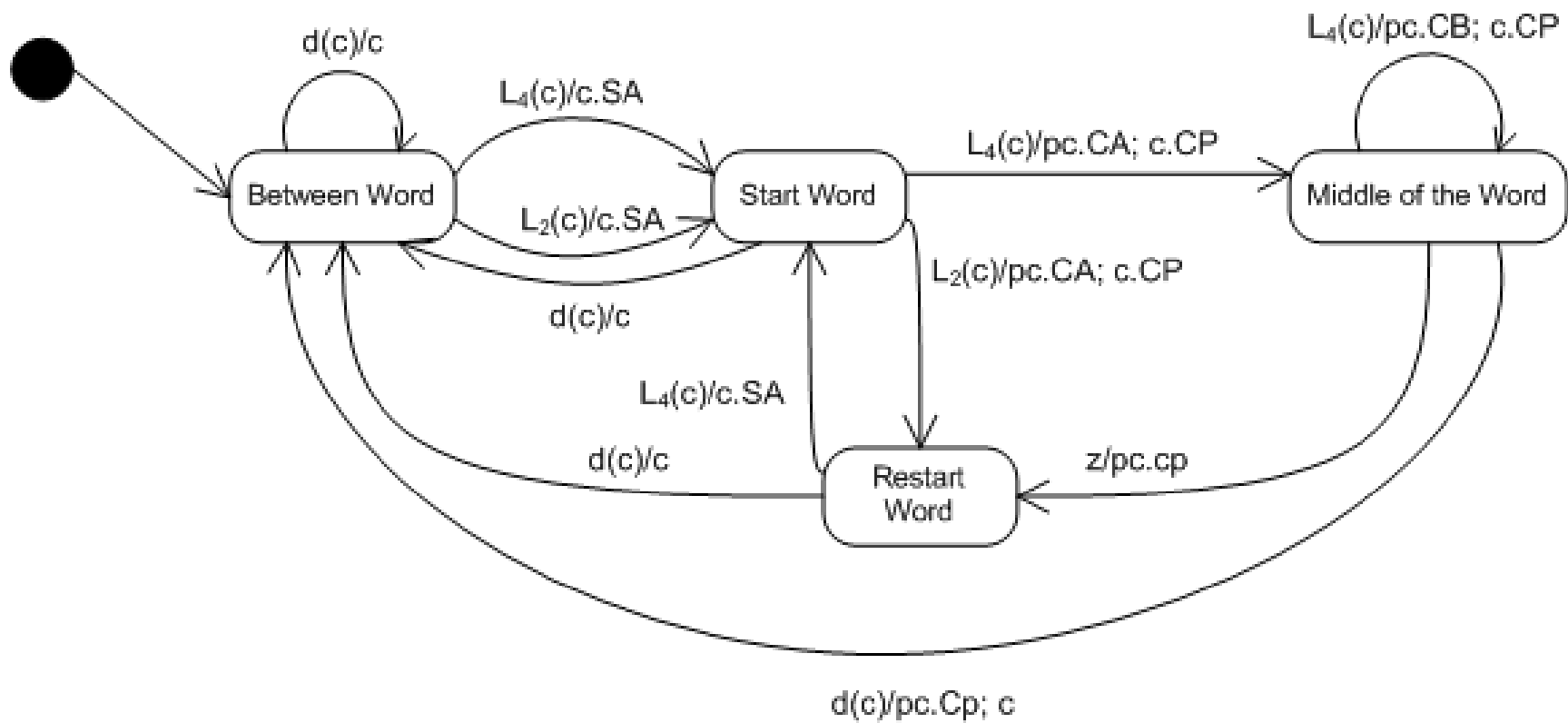


Sample: `_bblb_`

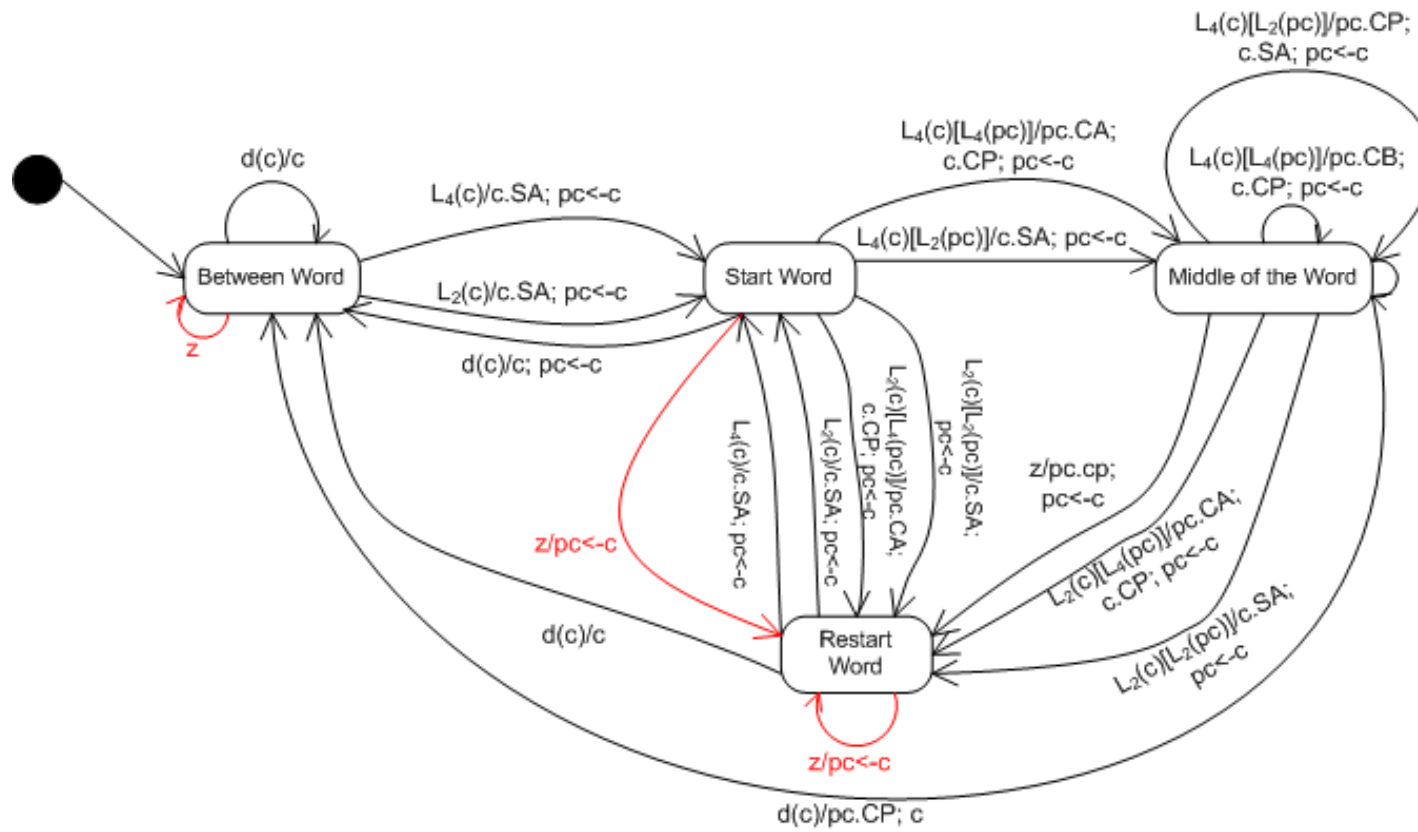




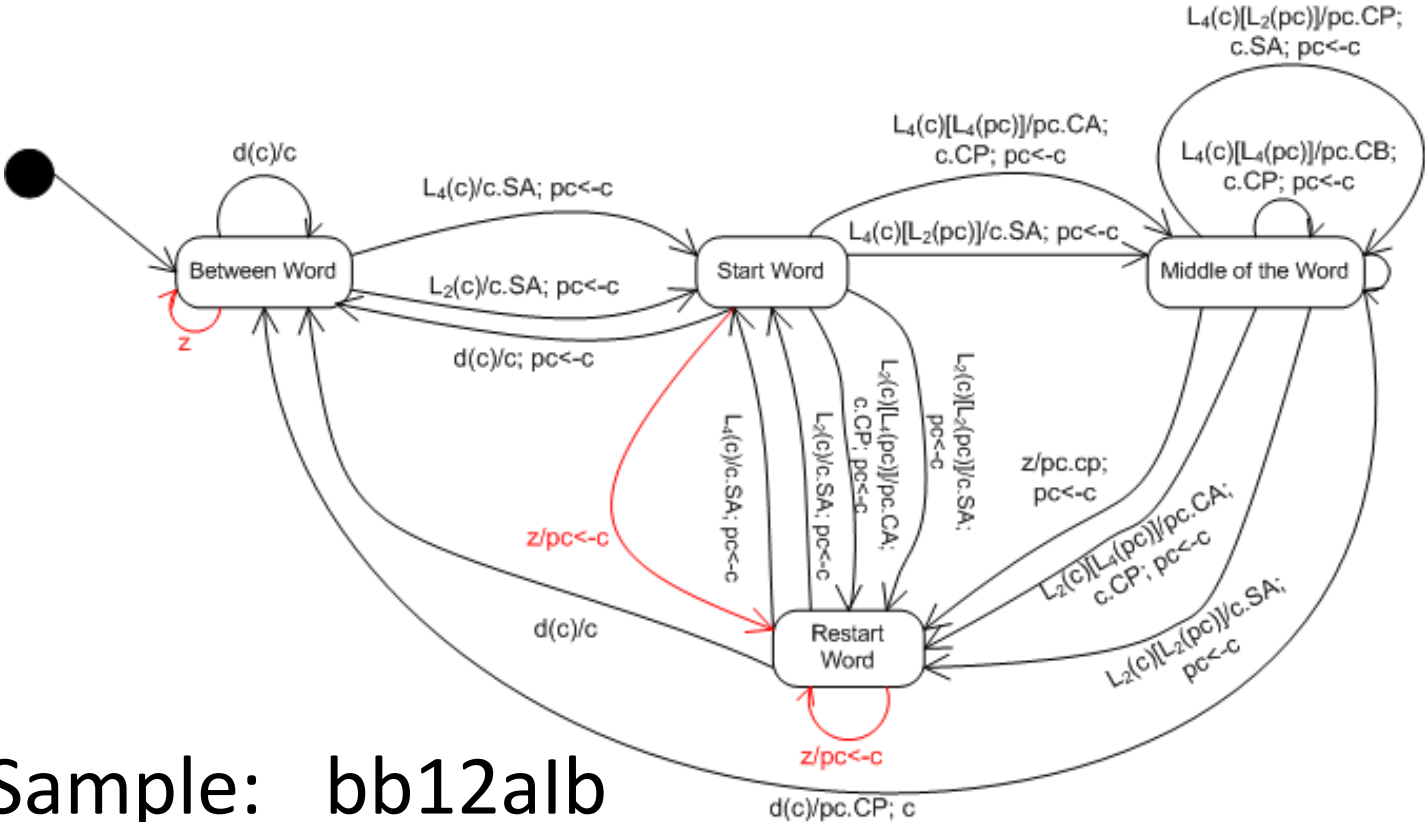
Sample: `_bab_`



VAR pc:char
 VAR c:char



VAR pc:char
VAR c:char



Sample: `_bb12alb_`