Arabic PropBank and Arabic VerbNet

Yahya Aseri

1- Introduction

- **Lexical resources**

- **English**
  - WordNet (✓)
  - VerbNet (✓)
  - PropBank (✓)
  - FrameNet (✓)

- semlink

- **Arabic :**
  - PropBank (✓)
  - VerbNet (✓)
  - WordNet (✓)
  - FrameNet (✗)

- WN and VN are mapped
- PB and VN (?)

1- Introduction

- The main questions:
- What are the challenges that computational linguists encountered when they tried to apply these ideas (PB or VN) on Arabic?
- What are the adaptations that have been made?

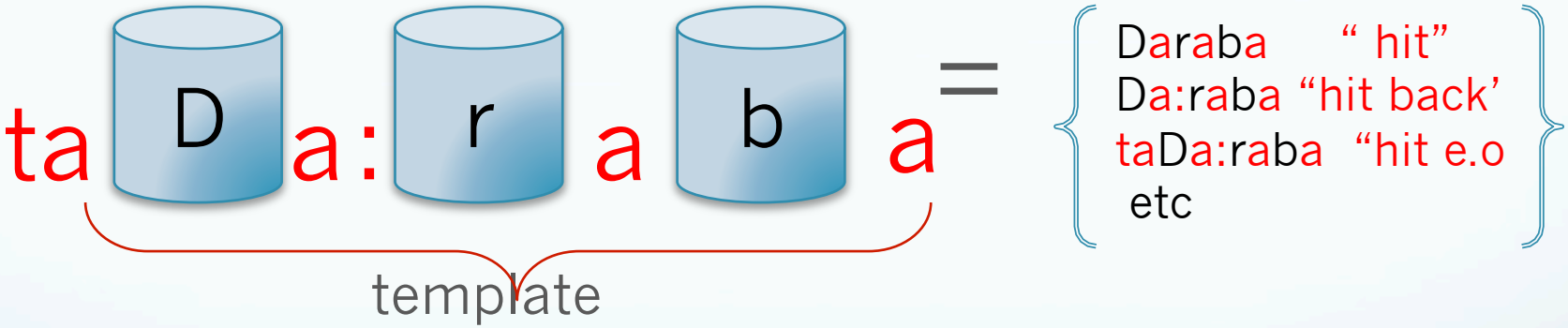
1-Introduction

- **Outlines:**
  1) A window on the Arabic Morphology
  2) Arabic PropBank (APB)
     - Paper: Martha and others (2008) *A Pilot Arabic Propbank*
  3) Arabic VerbNet (AVN)
     - Paper: Mousser (2010)*A Large Coverage Verb Taxonomy for Arabic*
  4) Mapping the APB to AVN.
  5) My final project

2- A window on the Arabic Morphology

- Arabic is a root- template language
  - This means two properties:

1) Given a root, different **lexical words** are produced from this root by variable templates (productivity).

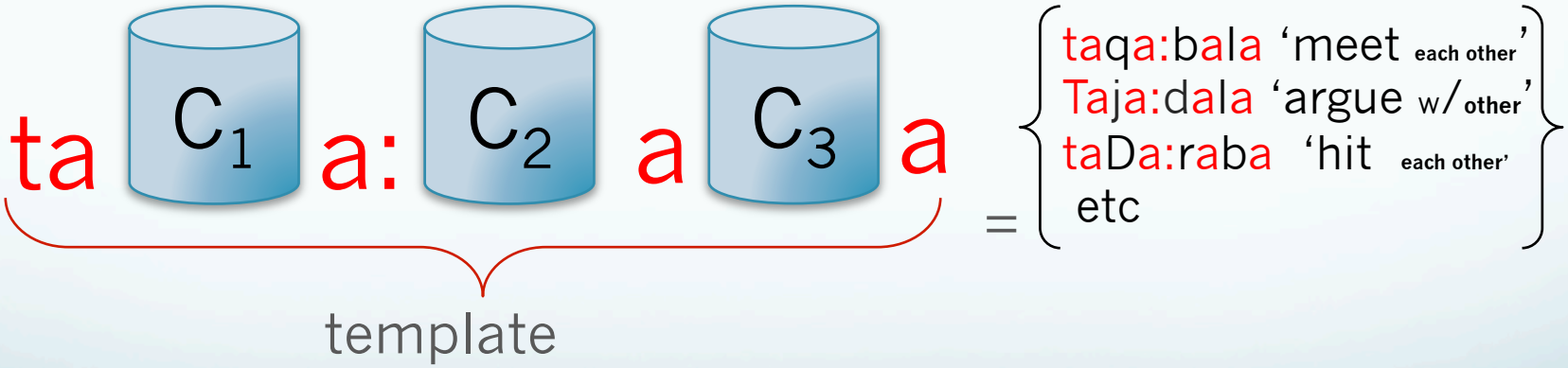


- The root consists of either 3 or 4 consonants.
  - 80% are three-consonant root verbs

2- A window on the Arabic Morphology

2) The second property:

Given a certain template, different **lexical words** across the lexicon are formed from variable roots (regularity).



2- A window on the Arabic Morphology

What does this mean?

2- A window on the Arabic Morphology

- Two lexical properties: Productivity and regularity are relevant to the morphological system.

productivity →

Regularity ↓

	Base form		derived form	
root	Base (4)	Deverbal nouns	Derived verb (11)	Deverbal nouns (8)
d/r/b	Daraba ‘hit’	Da:rib	Da:raba	maDrub
k/t/b	kataba ‘wrote’	Ka:tib	sa:taba	maktub
s/r/q	saraqa ‘stole’	sa:ariq	sa:raqa	masruq
w/S/l	waSala ‘arrived	wa:Sil	wa:sala	mawSul
k/s/r	kasara ‘ broke’	ka:sir	ka:sara	maksur
j/l/s	jalasa ‘ sit’	ja:lis	ja:lasa	majlus

2- A window on the Arabic Morphology

We care about semantics and syntax

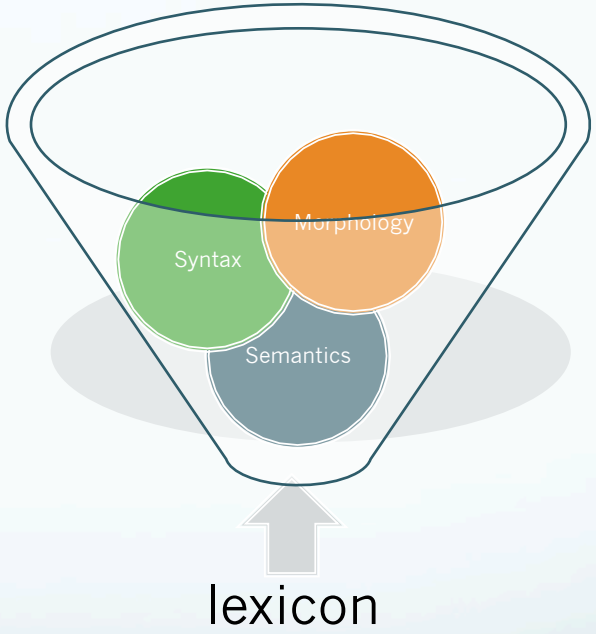
So;

Let's focus on the verb and ask about the morphological interaction with semantics and syntax

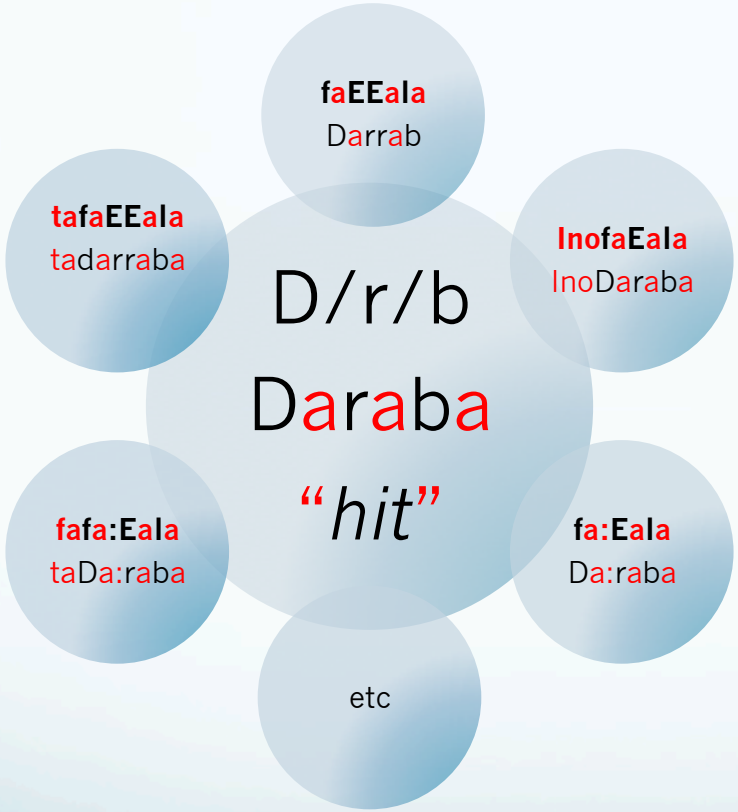
2- A window on the Arabic Morphology

- *Morphological interaction with syntax and semantics*
- *Questions:*
  - Are there syntactic and semantic correlations between derived verbs and their original ones/ base? (productivity)
  - For verbs that share the **same morphological template** across the lexicon, Do they also share some syntactic and semantic generalizations? (regularity)

  (my final project)



2- A window on the Arabic Morphology



- **Derivation and syntactic and semantic correlations:**

- The effects of derivational operations on the semantic and syntactic:
  1) changing the valence
  2) creating a new verb
  3) changing the aspect of the event.

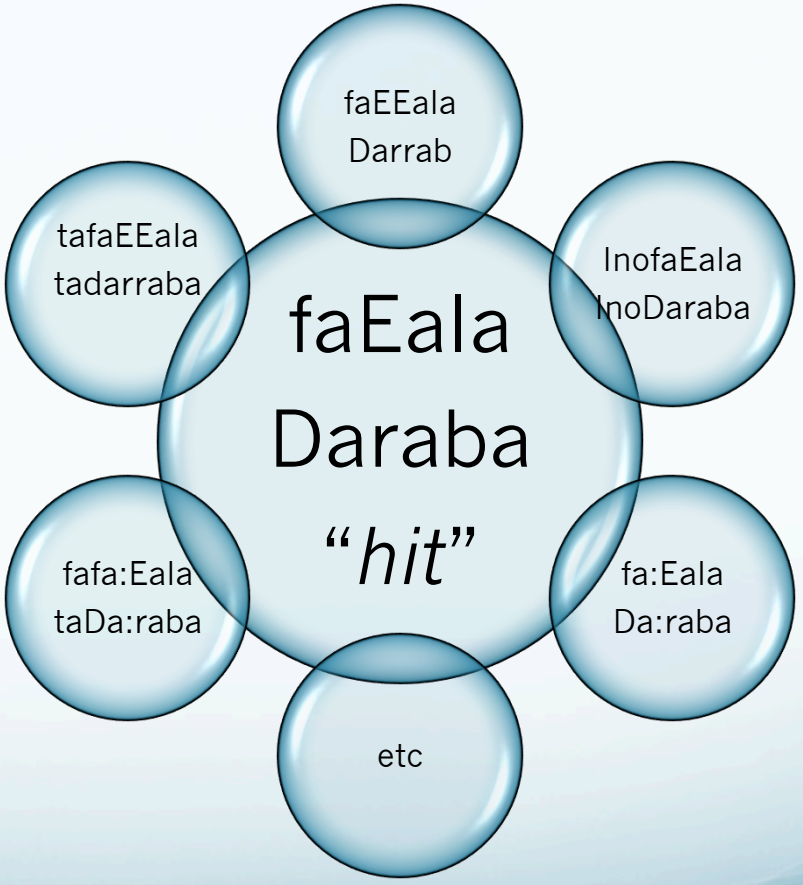
- In general, derived verbs often share the core meaning of the base verb.

2- A window on the Arabic Morphology

- **Classical dictionaries:**

- The root is the entry and all words derived from it are listed under this root:

(Root):
- Base verb
- Derived verbs
- Deverbal nouns
- Adjectives



2- A window on the Arabic Morphology

How did they deal with this issue in the APB and the AVN?

3- Arabic PropBank

- **Design:**
- The design is very similar in terms of the steps to the design steps taken for previous language:

- **Consists of two parts:**

1) Framefiles (lexicon)
   - frame for each predicate:
     - Predicate
     - Framesets
       - Description
       - Arguments
       - example
2) Annotated corpus:
   - syntactically parsed text provided by Arabic TB.

3- Arabic PropBank

- **Challenges and adaptation:**

  1) Lexical entry in the frame file:

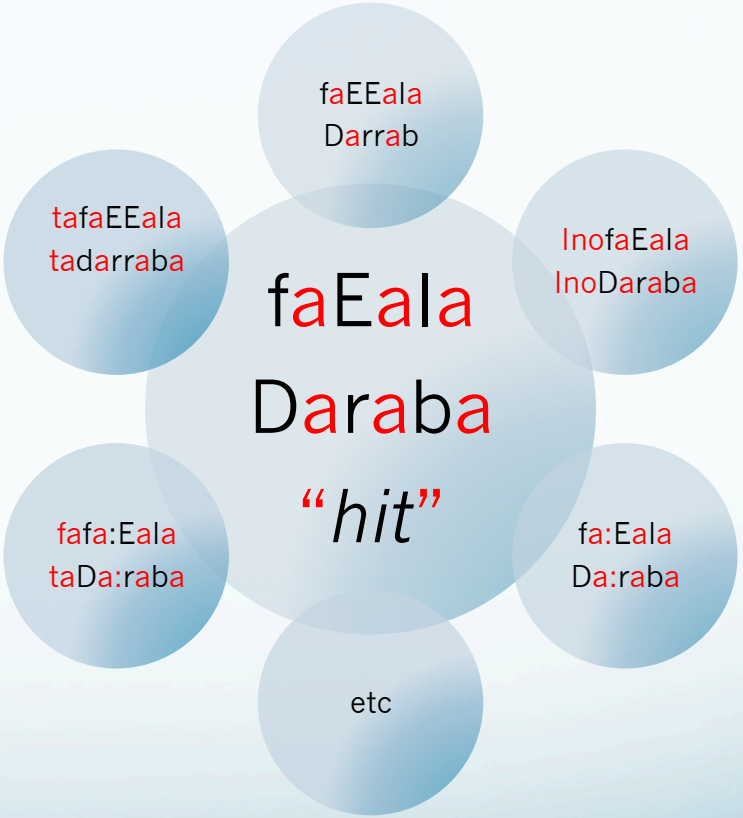
- Two approaches for choosing a lexical entry:

  1) Root approach:

  - High level of abstraction.
  - It does not meet the purpose of the PB

  2) Lemma approach:

  - good for capturing the argument structure for every single verb.
  - Risk of losing the connection among them.



3- Arabic PropBank

- **Challenges and adaptation:**

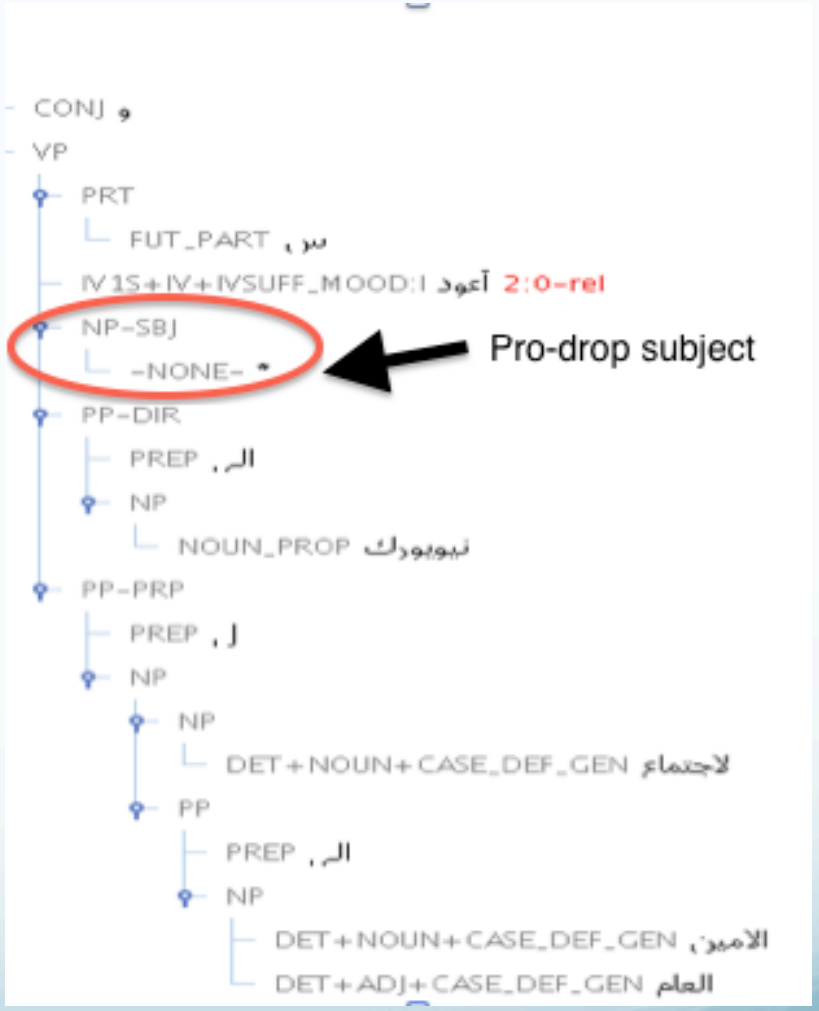
  2) Pro-drop subject:

- Arabic is a pro-drop subject language.
  - example:

سوف أعود إلى نيويورك

will | come back | to | New York

- *How do we tag this missing argument?*
  - Decision has been made by ATB
    - Creating a trace *



3- Arabic PropBank

- **Challenges and adaptation:**

  3) Passive Voice:

  - The distinction between passive voice and active is based on the short vowels combined with the root of the verb.
  - Arabic script is underspecified by nature for short vowels.
  - Example: (passive /active?)

فتح (فُتِحَ /فَتَحَ؟) الباب بالمفتاح

With the key | the door | open

- The decision has been made by ATB by specifying passive on the tree.

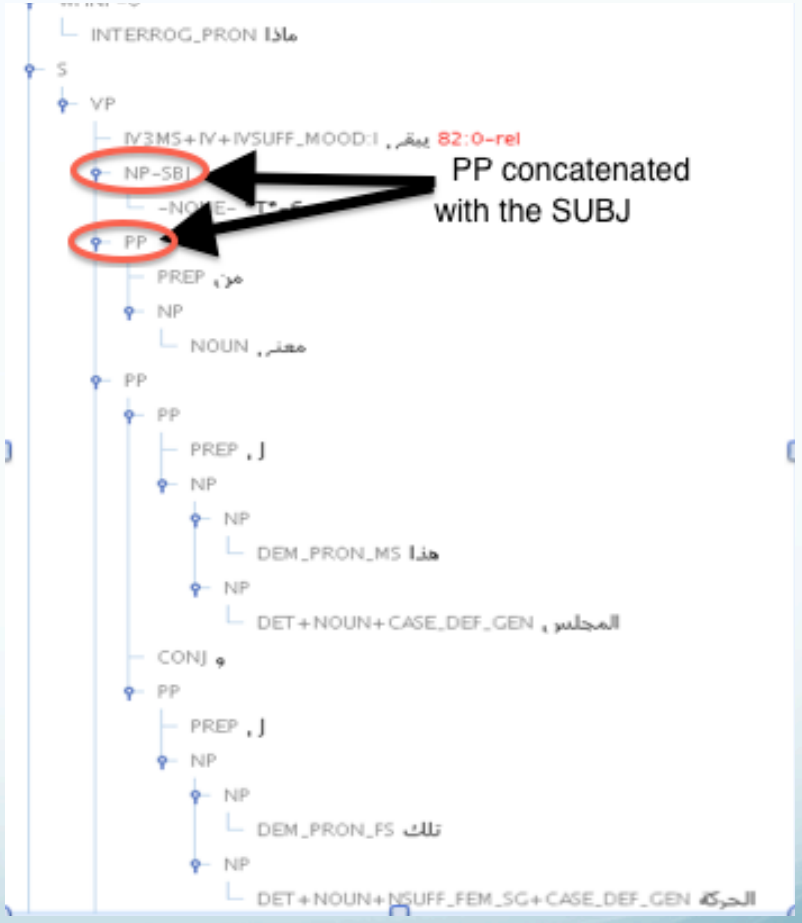
3- Arabic PropBank

- **Challenges and adaptation:**

  4) The Subject and PP issue:

- Frequently, Arabic allows a PP that substitutes an Argument.

- This PP is tagged by concatenating the two nodes



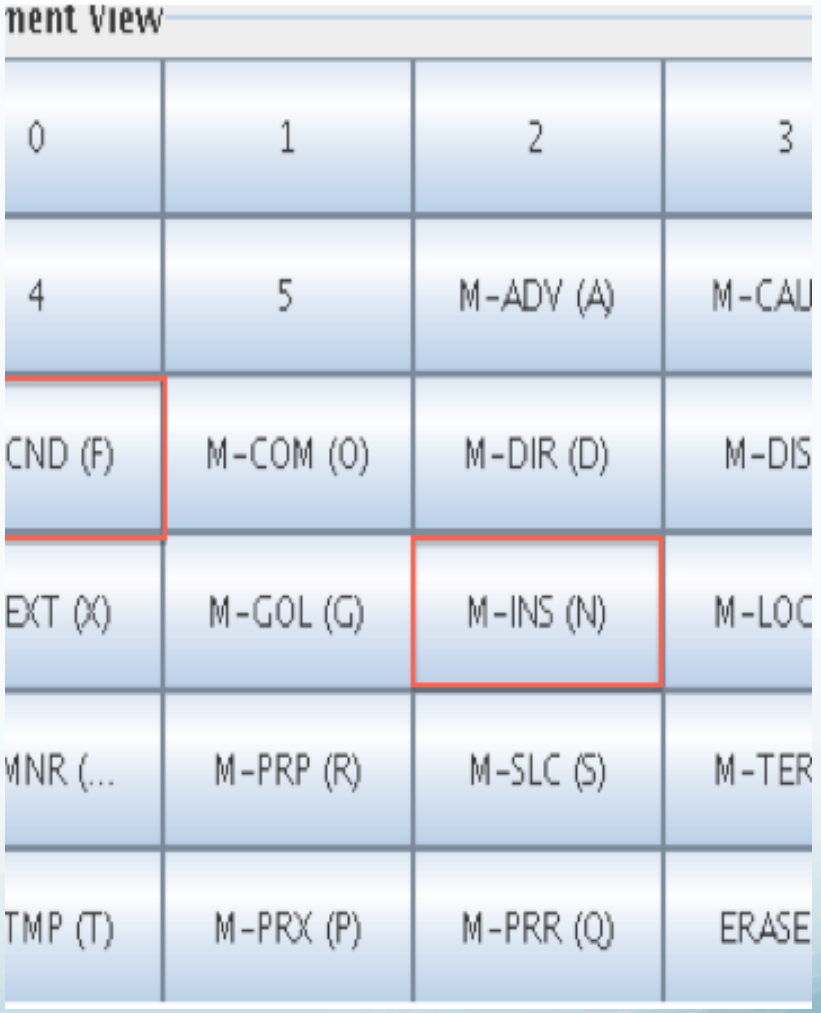
3-Arabic PropBank

- **Challenges and adaptation:**

  5) ARG-M:

  There is no deference between EPB and APB in terms of the ARGs, however, there are some differences between them in ARG-M:

English	Arabic
_	**CND**
_	**INS**
ADJ	_
PRD	_
DSP	_
MOD	_
NEG	_
REC	_



3- Arabic PropBank

- Current status:

Verbs			Nominals		
Files	**Pred**	**FSets**	**Files**	**Pred**	**FSets**
3220	**3220**	**4347**	**1940**	**1940**	**2241**

- Annotate:
  - Modern Standard Arabic (Newswires)
  - Some Dialects (Egyptian)

VerbNet

4- Arabic VerbNet

- Arabic VerbNet exploits Levin's verb classes and the procedure of developing English VN (2005).
- It has been built on the assumption that Verb Classes' idea can be transferred to Arabic with some adaptations.
- Basic Approach:
  - They used Levin's classes and some English novel classes.
  - Members of each class are translated to Arabic and expanded by using synonymy, hyponymy, hypernymy

4- Arabic VerbNet

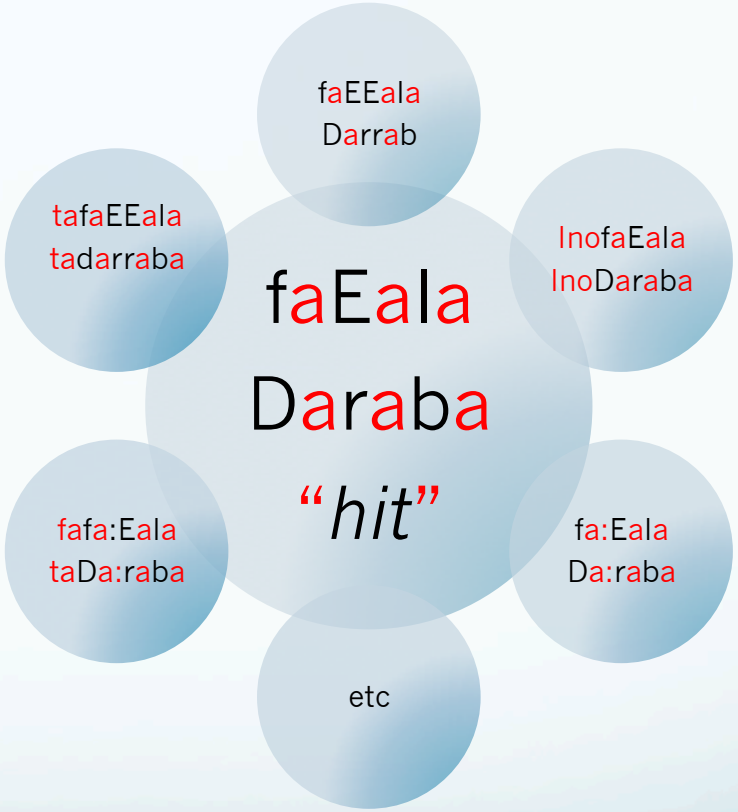
- **Arabic VN design**:
- similar to the EVN design
  - verbs are grouped into classes based on their semantic and syntactic behaviors

- Each class is organized in a hierarchal way:
  - Prototypical verb
  - Members
  - Thematic roles
  - Syntactic frame
  - Semantic description

4- Arabic VerbNet

- **Challenges and adaptation:**
- 1) derived verbs

- The effects of derivational operations on the semantic and syntactic could be
  1) changing the valence
  2) creating a new verb
  3) changing the aspect of the event.
- grouping verbs into classes results in losing the connection between derived verbs and the original form/base verb.



4- Arabic VerbNet

- adaptations:
  - Added to another class:

    If a derived verb is deferent from the original class, but it shares the properties of an existed class, it is added to that class.
  - subclass

    If a derived verb corresponds to the original class, but it adds additional semantic predicates separating it from the meaning of the original class, subclass is created
  - Sibling class:

    If a derived verb does not fit any class and the effect of derivation is only valence change. a sibling class is created and linked to the original one.

4- Arabic VerbNet

- **Challenges and adaptation:**

  2) diathesis alternations:

  - Preliminary study about diathesis alternations in Arabic was required to determine the deference between Arabic and English.
  - 65% of alternations in English are also available in Arabic.
  - Specificity of Arabic is due to the **morphological operations.**
  - **e.g :**
    - **amuse class:**
      - **English :** 6 frame (with two alternations)
      - **Arabic :** 4 frame (delete one and adds one)
  - Can we still talk about the same class when we add or delete some alternations in the target language?

    (sibling class)

4- Arabic VerbNet

- **Challenges and adaptation:**

  3) Class divergence

  - One class in English could be split into two groups of verbs/classes in Arabic:
    - Example:
  - class: *manner_of_speaking.*
    - **English**: *one class* **Arabic** *: two classes*

Group 1 **Agent- topic- recipient**	**Group 2** **Agent**
waswas ‘whisper’ *hashasa* ‘swish’ awh*aa* ‘reveal’ *hamasa* ‘whisper’ etc	*tamtama* ‘mumble’ *walwala* ‘make a howl’ *gaga* ‘growl’ *damdama* ‘burr’ etc

4- Arabic VerbNet

- **Challenges and adaptation:**

  4) Integrating two classes:

  - The properties of separating two class in English do not exist in Arabic
    - Class: *gobble* and *devour*
      - they have the same properties in Arabic

  5) Classes are not existed in Arabic:

    - Class: Debore

4- Arabic VerbNet

Challenges and adaptation:

6) Creating new classes.

- Levin's classes did not fit the event structure of some Arabic verbs.
  - *sara*: " walked during the night"
    - class of motion (like 'run'), but it adds a new predicate describing the property of walking during a specific time during the day or night.

5- Mapping the APB to the AVN

5- Mapping the APB to the AVN

back to the morphological regularity in the lexicon

5- Mapping the APB to the AVN

- Mapping based on the morphological template

productivity

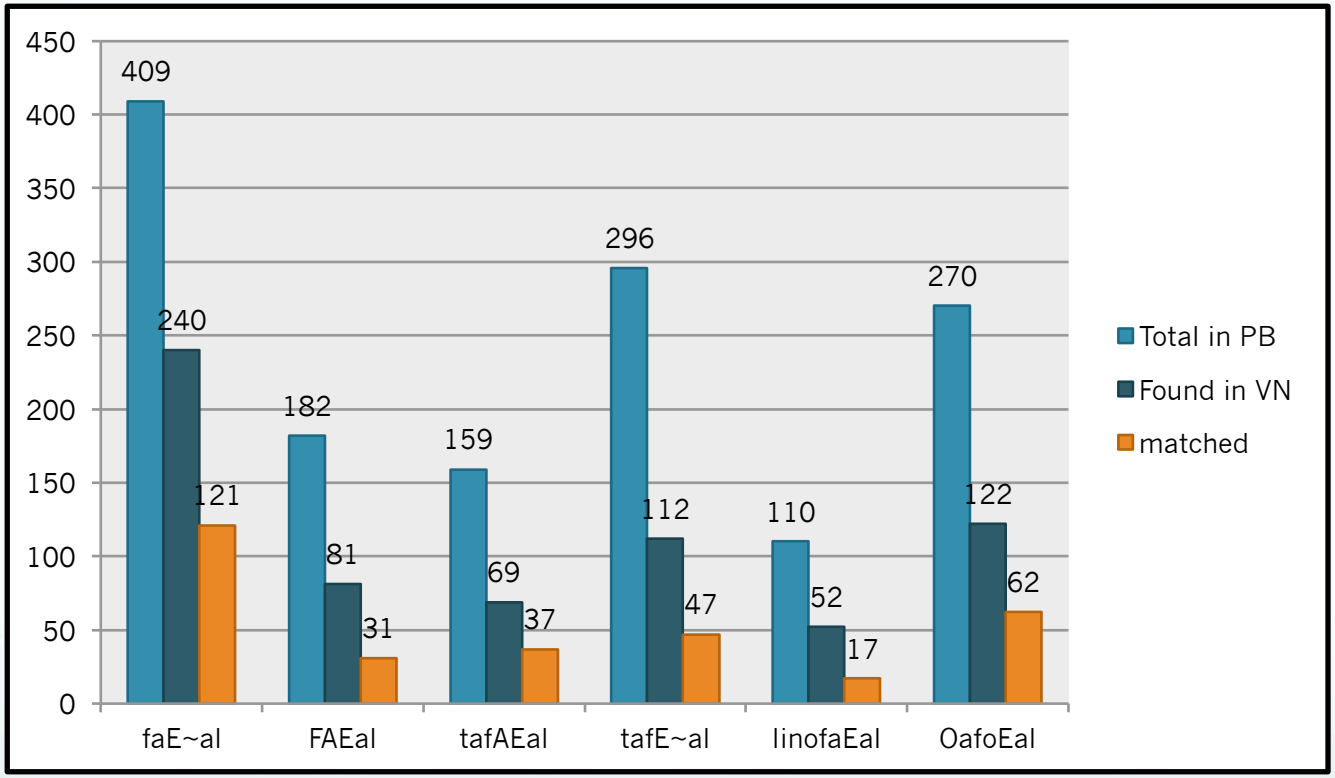
Regularity

	Base form		derived form	
root	base	Deverbal nouns	Derived verb	Deverbal nouns
d/r/b	Daraba ‘hit’	Da:rib	Da:raba	maDrub
k/t/b	kataba ‘wrote’	Ka:tib	sa:taba	maktub
s/r/q	saraqa ‘stole’	sa:ariq	sa:raqa	masruq
w/S/l	waSala ‘arrived	wa:Sil	wa:sala	mawSul
k/s/r	kasara ‘ broke’	ka:sir	ka:sara	maksur
j/l/s	jalasa ‘ sit’	ja:lis	ja:lasa	majlus

5- Mapping the APB to the AVN

- Mapping based on the morphological template:
- This approach allows us to :
  - map every single verb in the APB to the AVN
  - Also allows us to classify verbs into groups based on their morphological patterns, which allows us to:
    1) Compare the argument structure of derived verbs with the argument of their base verbs. (generate frame for PB)
    2) Investigate the argument structure of each template. (SRL)

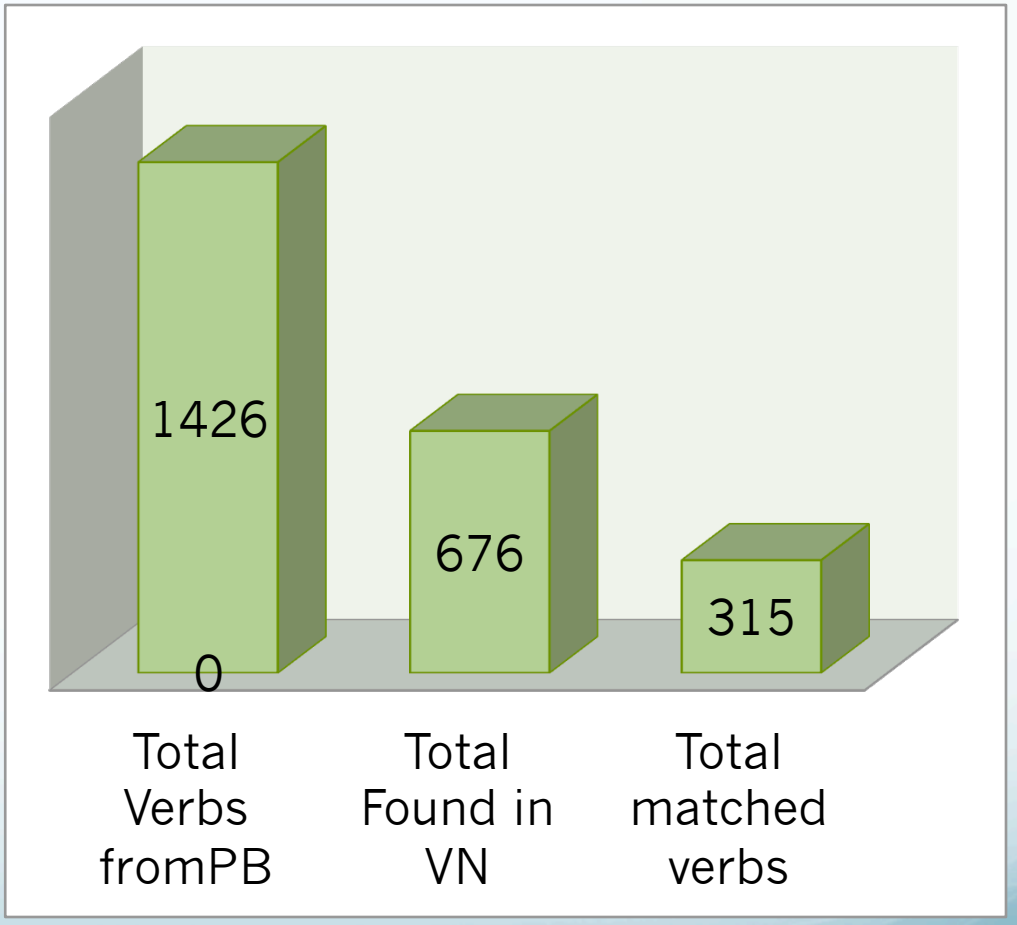
5- Mapping the APB to the AVN



Distribution of Verbs on the morphological templates

5- Mapping the APB to the AVN

- Total number of verbs extracted from APB is 1426
- Less than half was found in the AVN (47%)
- Less than half of the existed verbs in AVN was matched (46%)



6- My final Project

- *Towards Building a System for Semantic Role Labeling in Arabic.*

6- My final Project

back to this question:

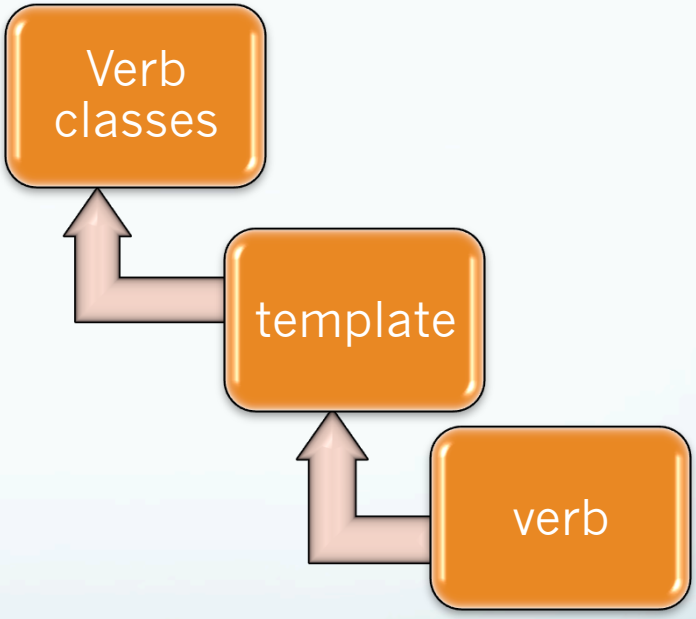
For verbs that share the same template across the lexicon (e.g. **$C_1aC_2C_2aIa$**), Do they share some syntactic and semantic generalizations?

6- My final Project

- **My assumption:**

Verbs generated by a certain template across the lexicon would share some semantic and syntactic behaviors.

- **The aims of this project:**
  - Introduce an approach for building a system for semantic role labeling in Arabic that takes into account the morphological interaction with syntax and semantics.
  - To do a pilot study on only one template(**$C_1aC_2C_2aIa$**) that is already mapped so I can test my assumption if it suggests a large-scale study.



Thank you

شكراً جزيلاً