VERB-NOUN COLLOCATION IN ENGLISH-ARABIC MT ENGINES: UTILIZING

VERB-NOUN COLLOCATION IN DISAMBIGUATING VERB POLYSEMY

By: Hussein Soori

Polysemy in Different Verb Patterns

-be aware of by using the power of sight. (see 2:770)	Subject +vt+ noun/pronoun +infinitive (phrase), (vp18A, T64), (C5 (ii))	-I saw him hit the cat. (Hornby,1975: 65)	رأى	رأيته يضرب القطة.
-understand; learn by search or inquiry or reflection. (see 4:771)	Subject + vt + that-clause, (vp9, T45), (B3 (i))	-I saw that he disliked the cat. . (Hornby, 1975: 65)	فهِمَ	فهمت أنه لا يحب القطة.
-call up a picture of; imagine. (see 10:771)	subject + vt + noun/ pronoun (DO) + *as/like/for + noun (phrase)* OR *a clause introduced by as if/ as though* (vp16B, T61), (C2 pr (ii))	-He saw himself as the saviour of this country. (OALDCE, s.v. see 10:771)	تصورَ	تصور نفسه منقذ هذا البلد.

+	T64	C5(ii)	(ii) Perception verb	Dynamic, Momentary event
+	T45	B3(i)	(i) Factual, Private	Stative, Private, Intellectual state
+	T61	C2 pr (ii)	(ii) Preposition verb, Obligatory (as)	Stative, Private, Intellectual state

Polysemy in the Same Verb Pattern

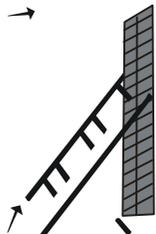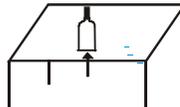|				
-Convey; transport. (run 15: 745)	subject + vt + noun/ pronoun (DO) + adverb (phrase) , (vp15A, T56), (C3)	-He ran me to the office. (OALDCE, s.v. run 15: 745)	اقلَّ	اقلَّني إلى المكتب.
-Cause to move (in a certain direction or into a certain place). (run 16: 745)	subject + vt + noun/ pronoun (DO) + adverb (phrase) , (vp15A, T56), (C3)	- He ran the car into the garage. (OALDCE, s.v. run 16: 745)	حرَّك	حرَّك السيارة إلى المرأب.
- Cause (a liquid, molten metal, etc.) to flow.(run 20: 745)	subject + vt + noun/ pronoun (DO) + adverb (phrase) , (vp15A, T56), (C3)	- He ran water into the bowl. (OALDCE, s.v. run 20: 745)	صبَّ	صبَّ الماء في الوعاء.

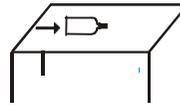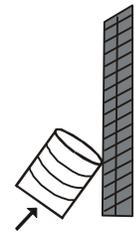|						
+	T56	C3	Causative verb of motion (in function)	Dynamic, Accomplishment	+human subject	+human object
+	T56	C3	Causative verb of motion (in function)	Dynamic, Accomplishment	+human subject	+concrete object
+	T56	C3	Causative verb of motion (in function)	Dynamic, Accomplishment	+human subject	+liquid object

The Problem

a. The painter *stood* the ladder *against* the wall.

أسنَد الدّهان السُلَّم إلى الجدار.

b. She *stood* the bottle *on* the table. (Ibid.)

وَضَعَت القنينة على الطاولة.

c. He *stood* the empty barrels *on* the floor. (Ibid.)

نَصبِ البراميل الفارغة على الأرض.

d. He *stood* the petrol tin *near* the fire. (Ibid.)

وضع علبة البنزين قرب النار.

e. She *stood* the toy *on* the table. (Ibid.)

ـ أوقَفَتْ الدمية على الطاولة

a-e above belong to vp15A, T56, C3, and carry one SL sense, i.e., "cause to be placed in an up right position" (Hornby and Cowie, 1980, s.v. stand 6) and the difference lies in the TL lexical choice.

Real World Knowledge?

Standing a ladder		Standing a bottle		Standing a barrel	
✔	✘	✔	✘	✔	✘

Real World Knowledge about Standing Objects

Frank Smadja (1993) Columbia University

 system >> Champollion - Xtract a statistical method

- Champollion's approach includes three stages:

1. Identify syntactically/semantically meaningful units in the source language.

2. Decide whether the units represent constituents or flexible word pairs.

3. Find matches in the target languages and rank them, assuming that the highest-ranked match for a given source-language collocation is its translation in the target language.

Prototype Main Components

```
Translation Engine
    ├── Parser
    ├── Lexicon
    └── Collocates Locater
```

Diagram of Collocates Locator: A statistical Approach

Arabic Corpus OR Online Open Source

Collocates, Collocates, Collocates
Collocates, Collocates, Collocates
Collocates, Collocates, Collocates
Collocates, Collocates, Collocates
Collocates, Collocates, Collocates
Collocates, Collocates, Collocates
Collocates, Collocates, Collocates
Collocates, Collocates, Collocates
Collocates, Collocates, Collocates
Collocates, Collocates, Collocates
Collocates, Collocates, Collocates
Collocates, Collocates, Collocates
Collocates, Collocates, Collocates
Collocates, Collocates, Collocates
Collocates, Collocates, Collocates
Collocates, Collocates, Collocates
Collocates, Collocates, Collocates
Collocates, Collocates, Collocates
Collocates, Collocates, Collocates
Collocates, Collocates, Collocates
Collocates, Collocates, Collocates
Collocates, Collocates, Collocates

Arabic Dictionary meanings

+

(TL Noun)
برميل

Dictionary

Meanings of the verb *stand* in Arabic:

أسند

وَضَعَت

نَصب

وضع

أوقَفَتْ

Word to the right

SL Verb   +   SL Noun
(collocate)   (base)
stand   ***barrel***

Word to the left

TL Verb   +   TL Noun
(collocate)   (base)
?   ***برميل***

نَصب

السُّلَّم
القنينة
علبة
الدمية

How many times the noun (*برميل*) occurred with the different verbs:

أوقَفَت	وضع	نَصب	وَضَعَت	أسند
15	4	**49**	2	23

- Stanford online parser

- Leeds University Arabic corpora

The man *put on* the shoes.	(ROOT
	  (S
	    (NP (DT The) (NN man))
	    (VP (VBD put)
	      (PP (IN on)
	        (NP (DT the) (NNS
	shoes))))
		det(man-2, The-1)
		nsubj(put-3, man-2)
		root(ROOT-0, put-3)
		prep(put-3, on-4)
		det(shoes-6, the-5)
		pobj(on-4, shoes-6)

The man *put on* the shirt.	(ROOT
	  (S
	    (NP (DT The) (NN man))
	    (VP (VBD put)
	      (PP (IN on)
	        (NP (DT the) (NN
	shirt))))
		det(man-2, The-1)
		nsubj(put-3, man-2)
		root(ROOT-0, put-3)
		prep(put-3, on-4)
		det(shirt-6, the-5)
		pobj(on-4, shirt-6)

The boy put on a cap.	(ROOT
	  (S
	    (NP (DT The) (NN boy))
	    (VP (VBD put)
	      (PP (IN on)
	        (NP (DT a) (NN
	cap)))
	    (. .)))
		det(boy-2, The-1)
		nsubj(put-3, boy-2)
		root(ROOT-0, put-3)
		prep(put-3, on-4)
		det(cap-6, a-5)
		pobj(on-4, cap-6)

Verb-Noun Training Data

SL Verb+ Noun Collocates in Sentences	Equivalent TL Verb-Nouns Collocates in Sentences (human translations)	Verb-Noun Syntactic Pattern	Lexical meaning from OALD (2010)
Noun as an Object			
1- The verb, *stand* - The painter *stood* the ladder against the wall. - The woman *stood* the bottle on the table. - The boy *stood* the empty barrels on the floor. - The man *stood* the petrol tin near the fire. - The girl *stood* the toy on the table.	أَسنَد الدّهان السُلّم على الجدار. وَضَعَت المرأة القنينة على الطاولة. نَصب الولد البراميل الفارغة على الأرض. وَضع الرجل علبة البنزين قُرب النار. أوقَفَتْ الفتاة الدمية على الطاولة.	verb + noun (object) + **prep.**	*stand*: Stand (verb) 3: to put sth/ sb in a vertical position somewhere.
2- The verb, *close*. - The director *closed* the deal. - The investigator *closed* the case.	أبرم المدير الصفقة. أغلق المحقق القضية.	verb + noun (object)	*close*: End (verb) 5: to end or make sth end.
3- The verb, *put on* - The man *put on* the shoes. - The man *put on* the shirt. - The boy *put on* a cap.	انتعل الرجل الحذاء. ارتدى الرجل القميص. ارتدى الولد كوفية.	verb + noun (object)	*put on*: Put sth on (verb) 1: to dress yourself in sth.
4- The verb *draw* - The waiter *drew* the cork out of the bottle.	نزع الجرسون الغطاء/ الفلينة عن	verb + noun (object) + **adv./ prep.**	*draw*: Pull sth/ sb (verb) 2: to move sth/ sb by pulling it or them gently.

- The doctor *drew* the chair up closer to the fire.	الزجاجة. سحب الدكتور الكرسي قريباً من النار.		
5- The verb *draw* - The policeman *drew* his gun. - The swordsman *drew* a sword.	سحب الشرطي مسدسه. استل المبارز سيفاً.	verb + noun (object)	*draw*: Weapon (verb) 6: to take out a weapon, such as a gun or a sward.
6- The verb *pass* - The parliament *has passed* the law. - The parliament *has passed* the proposal.	سن البرلمان القانون. مرر البرلمان المقترح.	verb + noun (object)	*pass*: Law/ proposal (verb) 14: to accept a proposal/ law by voting.
7- the verb, *pay* - The minister *paid* attention to the speech. - The director *paid* tribute to her contributions. - He *paid* a visit to the dermatologist last month.	أعار الوزير الانتباه للخطاب. أجزل المدير الثناء على إنجازاتها. زار طبيب الامراض الجلدية الشهر الماضي.	verb + noun (direct object) + to + noun verb + noun (direct object) + to + noun **verb + preposition + noun (*i.e.* a prepositional phrase)**	*pay*: (verb T) 6: used with some noun to show that you are giving or doing the thing mentioned. **The last case doesn't belong to the one verb category and should be put in a different category (paid a visit to) and must be put in a separate entry in the lexicon.**
8 - the verb, *win* - The councilor *won* the support for his proposal. - The activist won the admiration of many people for his battle against cancer.	كسب عضو المجلس الثقة لمقترحه. نال الناشط إعجاب كثير من الناس في معركته ضد السرطان.	verb + noun (direct object) + preposition + noun (*i.e.* a prepositional phrase) **verb + noun (direct object) + preposition + noun (*i.e.* a prepositional phrase) + preposition + noun (*i.e.* a prepositional phrase)**	*Win*: (verb T) 3: to achieve or get sth that you want, especially by your own efforts.
Noun as a Subject			
1- The verb, *drop*		**verb + noun (subject)/ + preposition + noun (*i.e.***	*drop*: Become weaker/ less (verb I,T) 4: to become or

English examples	Arabic	Pattern	Definition
- The price *dropped* dramatically. - The temperature *dropped* tonight. - The wind *dropped* tonight. - The singer's voice *dropped* suddenly. - The Dutch team *dropped* to the fifth place.	هبط السعر بشكل كبير. انخفضت درجة الحرارة الليلة . خفت الريح الليلة . انخفض صوت المغني فجأة. هبط الفريق الهولندي إلى المركز الخامس.	**a prepositional phrase)**	make sth weaker, lower or less.
2- The verb, *fall* - A sudden silence *fell* upon the room. - Darkness *fell* quickly tonight.	خيم سكون مفاجئ على الغرفة. اطبق الظلام مبكراً الليلة.	verb + noun (subject)	*fall*: Happen/ occur (verb I) 9: to come quickly and suddenly.
3- The verb, *fall* - My eye *fell on* a curious object. - A shadow fell across her face.	وقعت عيني على جسماً غريباً. سقط الظل على وجهها.	noun (subject) + verb + prep.	*Fall*: (verb I) 11: to move in a particular direction or come in a particular position.
4- the verb, *run* - The car *ran* off the road into a ditch. - A whisper *ran* through the crowd.	انحرفت السيارة عن الطريق إلى خندق. سرت همسة بين الحشد.	verb + noun (subject) - adv./prep.	*run*: Move somewhere (verb) 14: to move, especially quickly, in a particular direction.
Patterns not Included in the Training Data			
verb + noun (noun as a state)			
The man *flared up*.	استشاط الرجل غضباً.	verb + noun (noun as a state)	This pattern doesn't pause a problem because it only requires word for word translation.
noun + verb (names of sounds)			
-Dogs *bark* -Trees *rustle*	نباح الكلاب. حفيف الأشجار .	noun + verb (names of sounds)	These have one to one equivalents in both Arabic and English.

Leeds University Arabic Corpora

- *Buckwalter Arabic Corpus 1986-2003 - Tim Buckwalter – Written - 2.5 to 3 billion words - Lexicography Public resources on the Web*

- *Leuven Corpus (1990-2004) - Catholic University Leuven, Belgium*
Written and spoken - 3M words
(spoken: 700,000)

- *Arabic Newswire Corpus(1994) - University of Pennsylvania LDC*
Written - 80M words - Education and the development of technology
- *Agence France Presse, Xinhua News Agency, and Umma Press.*

- *CLARA (1997)- Charles University, Prague - Written - 50M words*
Lexicographic purposes - Periodicals, books, internet sources from 1975-present.

- *A parallel corpus of the Qur'an in English and Arabic*

- *An-Nahar Corpus (2001)- ELRA - Written - 140M words*
General research - An-Nahar newspaper (Lebanon)

- *Al-Hayat Corpus (2002)- ELRA - Written - 18.6M words - Language Engineering and Information Retrieval - Al-Hayat newspaper (Lebanon)*

- *Arabic Gigaword (2003) - University of Pennsylvania LDC*

- *Written - Around 400M - Natuarl language processing, information retrieval, language modelling - Agence France Presse, Al-Hayat news agency, An-Nahar news agency, Xinhua news agency.*

- E-A Parallel Corpus (2003)- University of Kuwait - Written
3M words - Teaching translation & lexicography - Publications from Kuwait National Council

- General Scientific Arabic Corpus (2004)- UMIST, UK - Written 1.6M words - Investigating Arabic compounds
http://www.kisr.edu.kw/science/

- Classical Arabic Corpus (CAC) (2004) UMIST, UK - Written - 5M words - Lexical analysis research
www.muhaddith.org and www.alwaraq.com

- Multilingual Corpus 2004 - UMIST, UK - Written - 10.7M words (Arabic 1M) - Translation - IT-specialized websites

- SOTETEL Corpus - SOTETEL-IT, Tunisia - Written - 8M words Lexicography - Literature, academic and journalistic material - Free

- Corpus of Contemporary Arabic (CCA) 2004 - University of Leeds, UK - Written and spoken - Around 1M words - TAFL - Websites and online magazines - Available online

- Arabic Blogs (2009)and a corpus builder application - Shereen Khoja, Pacific University, Oregon, USA - Written - 131,836 words

- Essex Arabic Summaries Corpus (EASC 1.0) - Mahmoud El-Haj, University of Essex, UK - Written - 153 Arabic articles and 765 human generated extractive summaries of the article.

- International Corpus of Arabic (ICA) 2008-2009 - Bibliotheca Alexandrina (BA)- Written - 100M words - General linguistic research - A wide range of sources from the Internet representing different Arabic regions.

Possible Problems: 1/…

Verb – Noun Collocation (object)	Verb – Noun Collocation (subject)
- The director *closed* the deal. أبرم المدير الصفقة.	- The policeman *drew* his gun. سحب الشرطي مسدسه.
- The investigator *closed* the case. أغلق المحقق القضية.	- The swordsman *drew* a sword استل المبارز سيفاً.

Possible Problems: 2/…

- The man *put on* the shoes.	انتعل الرجل الحذاء.
- The man *put on* the shirt.	ارتدى الرجل القميص.
- The boy *put on* a cap.	ارتدى الولد كوفية

SL – Human Translation

The director closed the deal.

أبرم المدير الصفقة.

The investigator closed the case.

أغلق المحقق القضية.

The price dropped dramatically.

هبط السعر بشكل كبير.

The temperature dropped tonight.-

انخفضت درجة الحرارة الليلة .

The wind dropped tonight.

خفت الريح الليلة .

The singer's voice dropped suddenly.

انخفض صوت المغني فجأة.

The Dutch team dropped to the fifth place.

هبط الفريق الهولندي إلى المركز الخامس.

SL - Google Translator at 11:54am on Tuesday, 27th November 2012

The director closed the deal.

‫- أغلقت مدير الصفقة.‬

The investigator closed the case.

‫- أغلقت المحقق في القضية.‬

The price dropped dramatically.

‫- انخفض سعر بشكل كبير.‬

The temperature dropped tonight.-

‫- انخفضت درجة الحرارة هذه الليلة.‬

The wind dropped tonight.

‫- انخفض الرياح هذه الليلة.‬

The singer's voice dropped suddenly.

‫- انخفض صوت المغني فجأة.‬

The Dutch team dropped to the fifth place.

‫- انخفض المنتخب الهولندي إلى المركز الخامس.‬