

شبكة الكلمات العربية « المنهجية والتطبيق »

د. المعتز بالله السعيد طه^{*}

الكلمات الدالة: شبكة الكلمات، الأنطولوجيا، المدونة اللغوية، صناعة المعجم.
- مُلخّص.

تعرض هذه الورقة مدخلاً إلى الشبكات اللغوية للكلمات WordNets باعتبارها تمثيلاً معرفياً واسعاً لجوانب اللغة الإنسانية. ويُقدّم الباحث - فيها - منهجاً مقترحاً لبناء شبكة للكلمات العربية Arabic Word Net انطلاقاً من الشبكات المعيارية المنجزة، مع مراعاة طبيعة اللغة العربية الاشتقاقية ونظامها المعجمي والدلالي - على مستوى المباني والمعاني. كما يعرض الباحث لجوانب الإفادة من شبكة الكلمات المنشودة في تطوير الصناعة المعجمية للعربية المعاصرة، مع العناية بالمعجمات المتخصصة.

١ - مُقدمة.

(* باحث في علم اللغة بكلية دار العلوم بجامعة القاهرة.

١.١ - شبكة الكلمات.

يُمكنُ تعريفُ شبكةِ الكلماتِ بأنَّها قاعدةُ بياناتٍ مُعْجَمِيَّةٌ Lexical Database؛ أو بعبارةٍ أُخرى: شبكةٌ من العلاقاتِ المُعْجَمِيَّةِ Lexical Relations، تُعالجُ فيها نُصوصُ اللُّغةِ بتعيينِ أقسامِ الكلامِ (POS) Parts of Speech، ثُمَّ تُصنَّفُ في مجموعاتٍ مُترابطةٍ من المُترادفاتِ Synonyms / Sets of Cognitive (Synsets) التي تُعبِّرُ عن مفاهيمٍ ثابتةٍ^(١)، وتتفقُ في معانيها الدلاليةِ - وإن اختلفت في المعاني السِّياقِيَّةِ؛ وترتبطُ هذه المُترادفاتِ فيما بينها بشبكةٍ من العلاقاتِ الدلاليةِ Semantic Relations.

٢.١ - الخلفية والإرهاصات.

١.٢.١ - شبكة كلمات برينستون PWN.

ظَهَرَتِ فِكْرَةُ شَبَكَةِ الكَلِمَاتِ فِي اللُّغَةِ الإنكليزيَّةِ لتكوُنَ أداةً لتطوِيرِ صناعةِ المُعْجَمِ Lexicography وتطبيقاتِ الذِّكاءِ الصُّنْعِيِّ Artificial Intelligence (AI) ومُعالِجَةِ اللُّغَاتِ الطَّبِيعِيَّةِ Natural Language Processing (NLP). وكانت البدايةُ الحَقِيقِيَّةُ لشبكاتِ الكلماتِ في عامِ ١٩٨٥م، إذ بدأَ العَمَلُ في بِناءِ وتطوِيرِ شَبَكَةِ كَامِلَةٍ لمُفْرَدَاتِ اللُّغَةِ الإنكليزيَّةِ في مُخْتَبَرِ العُلُومِ الإدراكيَّةِ بجامعةِ برينستون Princeton الأمريكيَّةِ^(٢)؛ وتَمَكَّنَ الفَرِيقُ الَّذِي

(1) Fellbaum, C., (1998). WordNet: an electronic lexical database. Cambridge, Mass: MIT Press.

(2) Fellbaum, C., (1998). Ibid.

أشرف عليه عالم النفس الأمريكي جورج ميلر George Miller من بناء الشبكة وتعبئتها يدويًا قبل إتاحتها على الشبكة العنكبوتية تحت مسمى «شبكة كلمات برينستون Princeton WordNet». وقد بلغ عدد الكلمات المفردة Unique words في هذه الشبكة 155287 كلمة، تُضمُّ الأقسام الكلامية: الاسم Noun = 117798 كلمة، والفعل Verb = 11529 كلمة، والصفة Adjective = 21479 كلمة، والظرف (الحال) Adverb = 4481 كلمة (PWN 3.0 - 2011).
ويبلغ عدد المجموعات الترادفية Synsets التي تُضمُّ كلمات الشبكة 117659 مجموعة (82115 للأسماء، 13767 للأفعال، 18156 للصفات، 3621 للظروف)، حيث تترابط كلمات كل مجموعة عن طريق شبكة موسعة من العلاقات الدلالية، نستطيع التمثيل لها بالعلاقات الموضحة في الجدول التالي⁽¹⁾:

المثال	المفهوم (التعبير الرياضي)	العلاقة
canine is a hypernym of dog	Y is a hypernym of X if every X is a (kind of) Y	الاشتمالية Hypernyms
dog is a hyponym of canine	Y is a hyponym of X if every Y is a (kind of) X	النوعية Hyponyms
building is a holonym of window	Y is a holonym of X if X is a part of Y	الكئية Holonym
window is a meronym of building	Y is a meronym of X if Y is a part of X	الجزئية Meronym

(3) Fellbaum, C., (1998). Ibid.

المثال	المفهوم (التعبير الرياضي)	العلاقة
to lisp is a troponym of to talk	the verb Y is a troponym of the verb X if the activity Y is doing X in some manner	التَّلازُمية Troponym (خاصة بالأفعال)
to sleep is entailed by to snore	the verb Y is entailed by X if by doing X you must be doing Y	الاستلزامية Entailment

الجدول 1: من العلاقات الدلالية في شبكة كلمات برينستون PWN

WordNet Search - 3.1
- WordNet home page - Glossary - Help

Word to search for:

Display Options:

Key: S = Show Synset (semantic relations), "W" = Show Word (lexical) relations
Display options for sense: (gloss), "an example sentence"

Noun

- **dic-tion-ary**, **lex-ic-on** is a reference book containing an alphabetical list of words with information about them.

الشكل 1: نموذج من شبكة كلمات برينستون للغة الإنكليزية «PWN 3.1 - 2011»

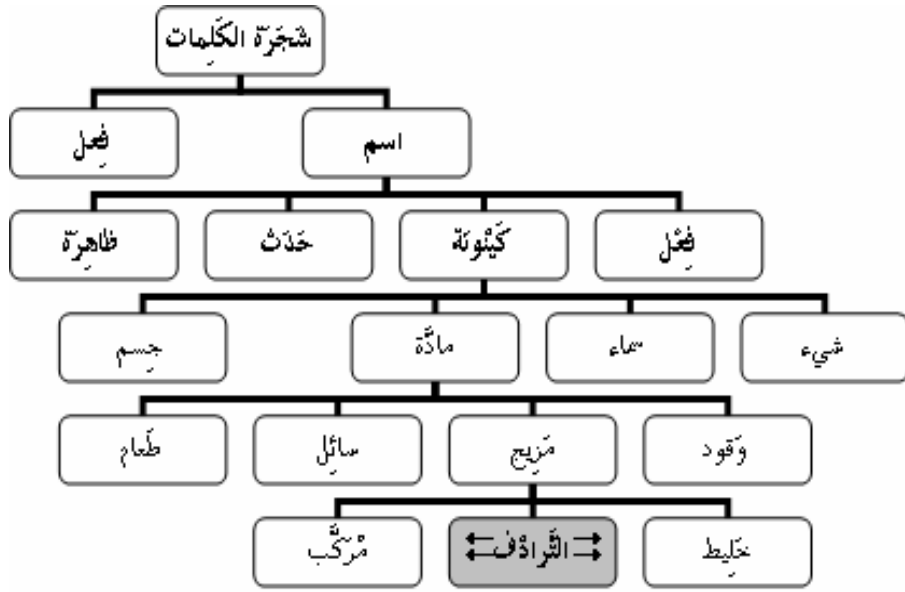
٢.٢.١ - شبكة الكلمات العربية AWN.

في المدة من ٢٠٠٥م إلى ٢٠٠٧م، سعى فريقٌ بحثيٌّ - في جامعات

برينستون وبرشلونة ومانشستر - لبناء شبكة للكلمات العربية على غرار شبكة برينستون. وأشرفت الباحثة الألمانية كريستيان فيلباوم Christiane Fellbaum - التي شاركت من قبل في بناء شبكة كلمات برينستون بقيادة جورج ميلر - على بناء شبكة الكلمات العربية AWN؛ إلا أن العمل توقف نتيجة نقص الموارد المتاحة لفريق العمل.

وتقوم الفكرة الأساسية لمشروع شبكة الكلمات العربية AWN على تزويد قاعدة بيانات الشبكة بمجموعة من المفردات «المفاهيم» المترجمة عن شبكة برينستون انطلاقاً من نظرية الحُقُول الدلالية Semantic fields حيث تشكل الشبكة في بناء شجري تُصنّف فيه المفردات إلى أسماء وأفعال، ويتفرّع عن كلٍّ منها مجموعة من الحُقُول التي تضم عدداً من المفردات الموزعة على حُقُول فرعية، وترتبط هذه المفردات فيما بينها بعلاقات دلالية⁽⁴⁾. ويبلغ عدد كلمات شبكة الكلمات العربية AWN 17561 كلمة مفردة؛ أما عدد المجموعات الترادفية Synsets التي تضم كلمات الشبكة فيبلغ ٧٨٢٢ مجموعة.

(4) Black, W., Elkateb, S., Rodriguez, H, Alkhalifa, M., Vossen, P., Pease, A. and Fellbaum, C., (2006). Introducing the Arabic WordNet Project, in Proceedings of the Third International WordNet Conference, Sojka, Choi, Fellbaum and Vossen eds.



الشّكل ٢: نموذج من شجرة الكلمات العربيّة في شبكة الكلمات العربيّة «AWN»
2.0 - 2007»

٣.١- شبكة الكلمات بين الشبكة الدلاليّة والأنطولوجيا.

تتكوّن شبكة الكلمات من مجموعة من الشبكات الدلاليّة Semantic Nets التي تربط بين المفاهيم باعتبارها شكلاً من أشكال التمثيل المعرفي؛ وفي الوقت ذاته يتكوّن من الارتباط بين مجموعة من شبكات الكلمات هيكل معرفي ضخم، يُستعمل في تطبيقات عديدة تتبع الذكاء الصنعيّ وعلوم المكتبات وصناعة المعجّات وتطبيقات الويب الدلاليّ Semantic web، ويُطلق عليه «الأنطولوجيا Ontology». وتمثّل الأنطولوجيا مجموعة من الكتل المعرفيّة التي تتلاحم عناصرها لتكوّن تمثيلاً لمجموعات منظمّة من المفاهيم المشتركة في حقول دلاليّة

مُحدّدة، مُترابطة بعلاقات دَلَالِيَّة^(١)، وتُستقى مادّتها من مُختلف فُرُوع المعرفة شريطة أن تُبرز واقع اللُّغة المُعيّنة.

وإذا كان الأصل في شبكات الكلمات أن تكون قواعد بياناتٍ مُعجميّة، تتلاحم عناصرها باستعمال العلاقات الدَلَالِيَّة، فإنّها - أي شبكات الكلمات - تُتمثّل مجموعةً من الأنطولوجيات الجزئية. وعلى هذا، فإنَّ وجودَ شبكةٍ معياريةٍ للكلمات العربية يفتح الطريقَ أمامَ بناء أنطولوجيا عربيةٍ كاملة؛ وهو مشروعٌ طموحٌ يعوقُ تحقيقه ضخامةُ المحتوى العربيّ وتناميه المُستمرّ، والغُموض في كثيرٍ من مفاهيم هذا المحتوى.



الشكل ٣: العلاقات بين المفاهيم (الشبكة الدَلَالِيَّة - شبكة الكلمات - الأنطولوجيا)

(5) Staab, S., Studer, R., (2004). Handbook on ontologies. Birkhäuser. And: Poli, R., Healy, M., Kameas, A., (2010). Theory and Applications of Ontology: Computer Applications. Springer.

٢- الإشكالية.

سارت شبكة الكلمات العربية AWN على منهج شبكة كلمات برينستون PWN في مادتها الأساسية. وإذا أردنا أن نتحرى الدقة، فيمكن القول إن شبكة الكلمات AWN ليست عربية بالمفهوم الدقيق، بل إنها ترجمة لبعض أجزاء شبكة كلمات برينستون PWN. وللووقوف على جوانب القصور التي يمكن أن تنتج عن مثل هذه الترجمة، فثمة بعض الحقائق التي تساعدنا على فهم طبيعة اللغات الإنسانية وأوجه التباين بينها، نحاول أن نجملها في النقاط التالية:

١. يستعمل الجنس البشري ما يربو على ستة آلاف وتسعمئة (٦٩.٩) لغة مختلفة في الحديث والكتابة، ولكل لغة من هذه اللغات مفرداتها التي لها نظائرها في اللغات الأخرى^(١).

٢. البيئة العربية هي التي صنعت مئات الكلمات لوصف الجمّل والجواد والسيف، وليس لذلك مثيل فيما عداها من لغات.

٣. في لغة جرينلاند القطبية ثلاث وستون كلمة بمعنى (ثلج)^(١).

٤. في لغة هاواي مئة وثمانين كلمة بمعنى (بطاطا حلوة)^(١).

(6) Finegan, E., (2011). Language: its structure and use. Thomson Wadsworth. 6th Edition. and Ginsburgh, V., Weber, S., (2011). How many languages do we need?: the economics of linguistic diversity. Princeton University Press.

(٧) السعيد (المعتر بالله): اللغة الإنسانية، دار الهاني، القاهرة، ٢٠١١.

٥. لا وجود للأرقام في قبائل بيرو البدائية إلا (من 1 إلى 4)، ويُعبّر عن غير ذلك من الأرقام بكلمة واحدة هي (tübāytē) (١).

٦. تختلف دلالة المفردات على المذكر والمؤنث من لغة إلى أخرى، ونلاحظ هذا بين العربية والألمانية - مثلاً - في كلمات مثل (قهوة Kaffee، طاولة Tisch) حيث تدل على مؤنث في العربية ومذكر في الألمانية، ومثلها (الباب Tür) التي تدل على مذكر في العربية ومؤنث في الألمانية.

إن طبيعة اللغات الإنسانية Human Languages تقتضي أن تختلف عناصر الأنظمة اللغوية وتتباين فيما بينها باختلاف الزمان والمكان؛ وهذا حال اللغة العربية، إذ تتمتع ببعض الخصائص التي تميزها عن غيرها من لغات الفصائل اللغوية الأخرى، بل وتُميزها - كذلك - عن بعض شقيقاتها من اللغات السامية. فهي لغة اشتقاقية، يقوم نظامها المعجمي على توليد الألفاظ والمعاني؛ وهي - كذلك - لغة مُعرّبة، تتغير أواخر كلماتها تبعاً لموقعها الإعرابي. وبعبارة أخرى تتسم اللغة العربية ببعض السمات التي تميزها عن غيرها من اللغات الإنسانية على مستوى المباني والمعاني.

فعلى مستوى المباني، تتشكّل ألفاظ اللغة العربية من أربعة وثلاثين فونياً قطعياً Segmental Phoneme؛ وينتج عن التداخل بين الفونيمات (المنطوقة)

(٨) السعيد (المعتز بالله): السابق.

(٩) السعيد (المعتز بالله): السابق.

تداخلُ بين الجرافيمات (المكتوبة)؛ وهذا ما يؤدي إلى تعيُّر شكل الحرف الواحد في الكلمة تبعاً لموضعه فيها من ناحية، وتعدُّد الألفاظ والمعاني المُسدلة عن المجموعاتِ الكتابيةِ المتَّفقةِ في حُرُوفها من ناحيةٍ أخرى.

المعنى	PoS	الكلمة	
كتب			
Write	فِعْل	Kataba	كَتَبَ
Writing	مَصْدَر	Katb	كُتِبَ
Books	جَمْع	Kotob	كُتُب
بل			
But	أداة	Bal	بَلْ
Wet	فِعْل	Balla	بَلَّ
Drencher	اسم	Ball	بَلٌّ

الجدول ٢: نموذج - التداخل الفونيمي والجرافيمي للأشكال الكتابية في العربية

أمَّا على مُستوى المعاني، فيبدو أنُّ البيئة العربية واضحةً في تعدُّد دلالة المفردة الواحدة من ناحية، وتعدُّد المفردات التي تُشيرُ إلى دلالةٍ واحدةٍ من ناحيةٍ أخرى. فمثالُ تعدُّد دلالة المفردة الواحدة: كلمة (حاجب) التي تُشيرُ إلى العديد من الدلالات، منها:

المقابل الإنجليزي	الدلالة العربية
Usher	مُسَاعِدَ المحكِّمة
Doorman	خازن الباب وحارسه
Brow	الشَّعْرُ النَّائِبُ فَوْقَ العَيْنِ
Hajeb	عَلَمٌ مُذَكَّرٌ (من أسماء الذُّكور)

الجدول ٣: نموذج - تعدُّد دلالة كلمة (حاجب) في العربية

ومثالُ تعدُّد المُفْرَدَاتِ الَّتِي تُشِيرُ إِلَى دَلَالَةٍ وَاحِدَةٍ: الكلمات (سَبْعُ / غَضَنَفَرُ / حَارِثُ / هِزْبَرُ / أُسَامَةُ / حَيْدَرُ / بَهَنَسُ / كَيْثُ / قَسُورَةُ / فِرْناسُ / بَاسِلُ / دِرْغَامُ / حَمْزَةُ / ...)، إذ تُشِيرُ جَمِيعُهَا إِلَى (الأسد) فِيمَا يُعْرَفُ بِظَاهِرَةِ (التَّرَادُفِ Synonym).

ووفقاً لمنهج الشبكة المنجزة AWN، لا وجود لهذه الكلمات في قاعدة البيانات المعجمية، إذ تُشِيرُ جَمِيعُهَا إِلَى المُقَابِلِ الإنكليزيّ "Lion"، وإن كانت البيئة الإنكليزية، الَّتِي وُجِدَ فِيهَا الأَسْدُ مُنْذُ القِدَمِ، تَسْمَحُ بِالتَّمْيِيزِ بَيْنَ دَلَالَاتِ الكَلِمَةِ، حَيْثُ يُشَارُ إِلَى المُذَكَّرِ (أسد = Lion)، وإلى المُؤنَّثِ (كَبُوءَةٌ = Lioness) وإلى الصَّغِيرِ (شِبَلُ = Cub)؛ وهو أمرٌ لا نجدُه في كلمةٍ مُسْتَمَدَّةٍ مِنَ البيئَةِ العربيَّةِ، مِثْلَ كَلِمَةِ (جَمَل) الَّتِي يُشَارُ إِلَى دَلَالَتِهَا الثَّلَاثِ (المُذَكَّرِ «جَمَل»، والمُؤنَّثِ «نَاقَة»، والصَّغِيرِ «حُور») بِالمُقَابِلِ الإنكليزيّ «Camel».

تكمُنُ إشكاليّة التّرجمة إذن في أمرين:

١. اعتبار العربيّة لغةً لصقيّة، وهو ما يتنافى مع طبيعة مبانيها.

٢. إغفال أثر البيّنة في اللّغة، وهو ما يتنافى مع طبيعة معانيها.

٣- المنهجية المقترحة.

في ضوء ما ذكرناه حول مفهوم شبكة الكلمات وطبيعة الشّبكات المنجزة، يُمكن القولُ إنّهُ ليست لدينا - حتى الآن - شبكةٌ لكلمات اللّغة العربيّة بالمفهوم الدّقيق لشبكات الكلمات؛ ولذا، تقترحُ الدّراسةُ بناءً شبكةً للكلمات العربيّة انطلاقاً من المعايير المنهجية التي سارت عليها شبكة كلمات برينستون PWN والشّبكات اللّغويّة للكلمات التي أنجزت تحت مظلة شبكة الكلمات العالميّة Euro WordNet، Global WordNet، مثل: شبكة الكلمات الأوربيّة Euro WordNet (EWN) (١)، وشبكة الكلمات الحرّة للغة الفرنسيّة WordNet Libre du Français (WOLF) (٢)، والأنطولوجيا المعياريّة الموحّدة - سومو Suggested Upper Merged Ontology (SUMO) (٣)، وأنطولوجيا ياجو Yet Another Great Ontology (YAGO) (٤)، وغيرها.

ويُفترَضُ في مُحتوى شبكة الكلمات العربيّة المنشودة أن يكونَ مناسباً

(10) <http://www.illc.uva.nl/EuroWordNet/>.

(11) <https://gforge.inria.fr/projects/wolf/>.

(12) <http://www.ontologyportal.org/>.

(13) <http://www.mpi-inf.mpg.de/yago-naga/yago/>.

للمنطق اللُّغَوِيّ ومُوافقاً لطبيعة اللُّغة العربيّة الاشتقاقية ونظامها المعجمي، كما يُفترَضُ أن يُظهرَ صورةً للعربيّة المعاصرة على مُستوى المباني والمعاني، بما يتناسبُ مع الهدف من شبكات الكلمات في صناعة المُعْجَمَات اللُّغَوِيَّة وتطبيقات الويب الدَّلاليّ من ناحية، وبما يفتحُ الطَّرِيقَ أمامَ بناء الأنطولوجيا العربيّة الكاملة من ناحيةٍ أُخرى.

وإذا كانَ الأصلُ في شبكة الكلمات WordNet أن تكونَ قاعدةَ بياناتٍ مُعْجَمِيَّة Lexical database، أو بعبارةٍ أُخرى: أن تكونَ شبكةً من العلاقات اللُّغَوِيَّة المُتداخلة، لا تقتصر مادّتها على العلاقات الدَّلاليَّة بين المُفردات، وإنّما تمتدُّ لتشملَ العلاقات النَّاتجة عن مستويات التَّحليل اللُّغَوِيّ الَّتِي تتكوَّنُ عنها المُعْجَمَاتُ اللُّغَوِيَّة، فالدراسةُ تقترحُ ثلاثَ مراحلَ لبناء شبكةٍ لكلمات اللُّغة العربيّة، هي (الإدخال «التعبئة»، والإثراء «الإغناء»، والرِّبْطُ بين البيانات). وبيانُ هذه المراحل فيما يلي.

١.٣ - المرحلة ١: إدخال البيانات.

تعتمدُ هذه المرحلةُ على الموارد اللُّغَوِيَّة الأساسيَّة للشبكة المنشودة، وتمثّلُ في:

١.١.٣ - المُدَوَّنة اللُّغَوِيَّة Linguistic Corpus.

هي المصدر الأساسيّ لمُفردات شبكة الكلمات؛ وتمثّلُ المُدَوَّنة اللُّغَوِيَّة - عموماً - مجموعةً من نُصوص اللُّغة الَّتِي يمكنُ التَّعاملُ معها آلياً والتَّحكُّمُ في بياناتها ومُدخالاتها - إضافةً أو حذفاً أو تعديلاً - بالاستفادة من قواعدِ بياناتٍ

صُمِّمَتْ لتكونَ قادرةً على التَّعامُلِ مع هذه النُّصوص، حيثُ تُمَثِّلُ هذه القواعدُ مخزناً كبيراً للُّغة، يُرجَعُ إليه وقتَ الحاجة ويَتَحَمَّلُ أيَّ قدرٍ من النُّصوص التي يُمكن أن تُضَافَ إلى المادَّةِ الأساسيّةِ للمُدَوَّنَةِ اللُّغويَّةِ مُستقبلاً^(١).

ويُفترَضُ في المُدَوَّنَةِ اللُّغويَّةِ المُستعملة في تعبئة شبكة الكلمات العربيَّة أن تُضمَّ نُصوصاً من العربيَّةِ الفُصحى المُعاصرة، باعتبارِ هذا المُستوى (الفُصحى المُعاصرة) تمثيلاً لواقع اللُّغة العربيَّة (على المُستويين الجُغرافيِّ والمُوضوعيِّ) من ناحية، ومُعَبِّراً عن الاستعمال اللُّغويِّ للعربيَّة بين مُستعمليها من ناحيةٍ أُخرى.

٢.١.٣ - مسرَدُ المُفردات / الألفاظ Glossary of Words.

المُفردة (أو الكلمة) - في اصطلاح اللُّغويِّين - صيغةٌ ذات وظيفة لُغويَّة مُعيَّنة في تركيب الجُملة، تقوم بِدَوْر وحدة من وحدات المُعجم، وتصلح لأن تُفرد أو تُحذف أو تُحشى [بزيادة السَّوابق واللَّواحق والحشايا]، أو يُغيَّر موضعها، أو يُستبدل بها غيرها في السِّياق، وترجع مادَّتها غالباً إلى أصولٍ ثلاثة [١٣].
ويُعتَبَرُ مسرَدُ المُفردات مورداً لُغويّاً فرعيّاً، إذ يستمدُّ مداخله ووحداته من المُدَوَّنَةِ اللُّغويَّة بعد فهرسة نُصوصها آلياً. ولكن كيف تُصنَّفُ المُفردات / الكلمات في شبكة الكلمات العربيَّة المنشودة؟!

لقد صَنَّفَ النُّحاةُ القُدَماءُ الكلامَ العَرَبِيَّ إلى مجموعةٍ من المُكوِّنات /

(١٤) السَّعيد (المُعزَّز بالله): مُدَوَّنَةُ مُعجم تاريخيٍّ للُّغة العربيَّة «مُعالِجة لُغويَّة حاسوبية»، رسالة

العناصر التي تُعرفُ بأقسام الكلام PoS. وقسموها إلى ثلاثة أقسام، هي: الاسم، والفعل، والأداة (حرف له معنى). ويحيدُ بعض اللغويين المحدثين عن هذا التقسيم، حيث يتجه الدكتور إبراهيم أنيس (١٩٠٦ - ١٩٧٧) إلى تقسيم أجزاء الكلام تقسيماً رباعياً (الاسم، والفعل، والحرف، والضمير) (١)، على حين يتجه الدكتور تمام حسّان (١٩١٨ - ٢٠١١) وطائفة من الباحثين إلى التقسيم السباعي (الاسم، والصفة، والفعل، والضمير، والخالفة، والظرف، والأداة) (٢).

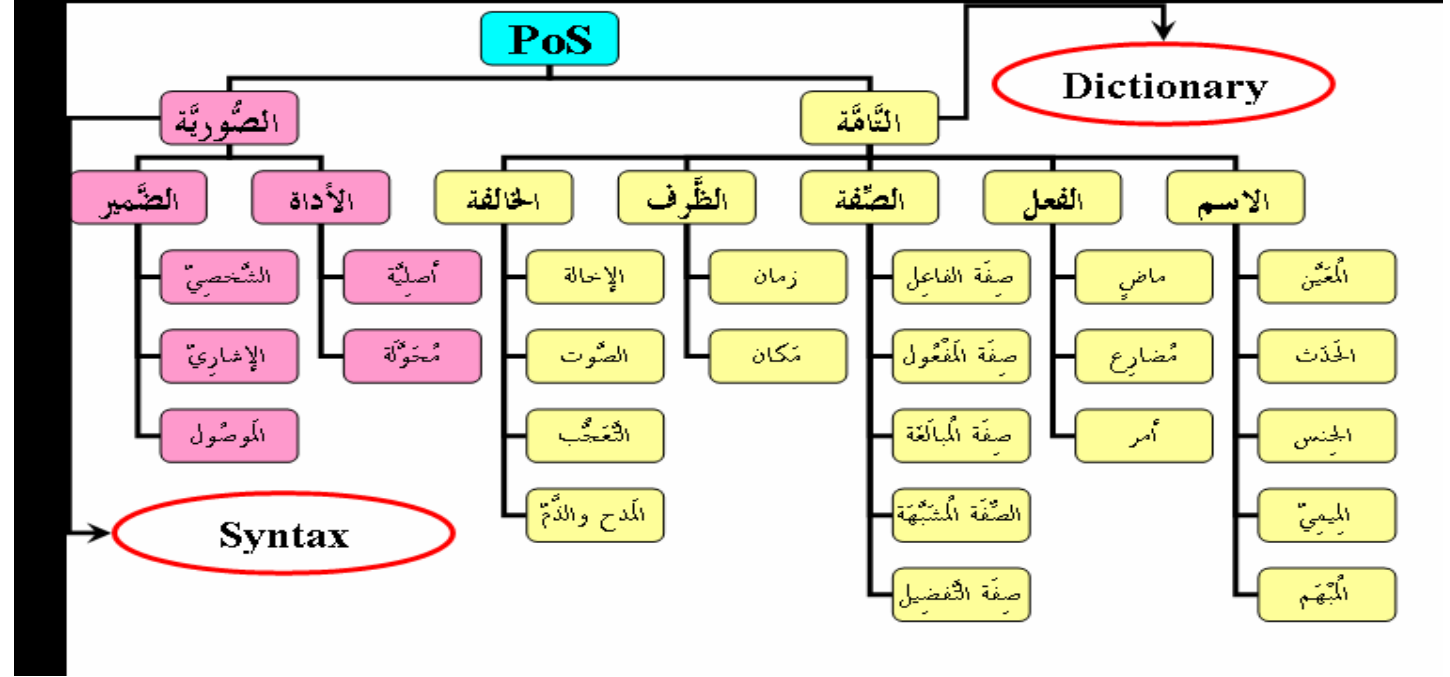
وتُصنّفُ الكلمات - عموماً - إلى كلمات تامّة Full words وكلمات صوريّة Form words؛ فالكلمات التامّة هي التي تحمل معنى تاماً في ذاتها - ومصدرها المعجم (Dictionary)، ومثّلها - وفق التقسيم السباعي: الأسماء (كتاب، قلم، مدرّسة، ...)، والأفعال (كتب، يفهم، يُغادر، ...) والصفات (أخضر، هادئ، سقيم، ...)، والظروف (تحت، أسفل، خلف، ...)، والخوالب (نعم، بس، حبذا، ...)؛ أمّا الكلمات الصوريّة فهي التي لا يُعرف معناها إلاّ بالنظر إلى الكلمات المجاورة لها، أو الجملة المرتبطة بها - ومصدرها علم التّركيب (Syntax)، ومثّلها: الأدوات (من، يا، ما، هل، ...)، والضمائر (نحن،

(١٥) أنيس (إبراهيم): من أسرار اللّغة، مكتبة الأنجلو المصريّة، القاهرة، ط٤، ١٩٧٢م.

(١٦) حسّان (تمام): اللّغة العربيّة "معناها ومبناها"، الهيئة المصريّة العامّة للكتاب، القاهرة، ط٢،

أنت، هي، هم، ...).

وإذا كان الهدف الأساسي من بناء شبكة الكلمات العربية أن تكون قاعدة بياناتٍ مُعجمية، وأداةً مُساعدةً في تقنيات الأنطولوجيا العربية والويب الدلالي، فإننا لن نكون بحاجةٍ إلى تزويد الشبكة المنشودة بقوائم الكلمات الصورية (الأدوات والضمائر)، إذ الأولى أن يُكتفى بالكلمات التامة التي تحمل معنى كاملاً يدلُّ عليها. وإذا كانت شبكة كلمات برينستون PWN قد صنّفت الكلام إلى أربعة أقسام، هي: الاسم والفعل والصفة والظرف، وحاكتها شبكة الكلمات العربيةAWN مُكتفيةً بتقسيم الكلام إلى الاسم والفعل، فإن الدراسة تقترح تقسيم الكلام إلى خمسة أقسامٍ مُستمدّةٍ من التقسيم السباعي العربي، هي: الاسم والفعل والصفة والظرف والخالفة.



الشكل ٤: الكلمات التامة والصورية في الكلام العربي - وفق التصنيف السباعي

٣.١.٣ - مسرد التعبيرات الاصطلاحية Idioms Glossary.

تُشير الكلمات المفردة إلى معانٍ مُعجمية حرفية أو مجازية؛ وتتعدّد هذه المعاني بتعدّد التنويعات السياقية وأساليب التعبير؛ وعلى سبيل المثال: تتعدّد المعاني المُعجمية لكلمة (عَيْن)؛ فقد تعني «عضو الإبصار»، وقد يُرادُ بها «ينبوع الماء في الأرض»، وترمي - مجازًا - إلى «الجناسوس»، وغيرها من المعاني التي تتضح من السياق.

أما التعبير الاصطلاحي - الذي ينتج عن تجمّع من الكلمات - فيُشير إلى معنى مجازي ثابت؛ ومن أمثله في العربية: التعبير (يغرق في شبر ماء) ومعناه (قليل الحيلة)، والتعبير (ضربَ كفاً بكفّ / ضربَ أحماساً لأسداس) ومعناه (تخيّر).

وإثراءً لمحتوى شبكة الكلمات العربية المنشودة، تقترح الدراسة تزويدها بمسردٍ للتعبيرات الاصطلاحية المستعملة في العربية المعاصرة، بحيث تُستمد مادّتها من المدونة اللغوية. ويُفترض أن يُمثّل هذا المسردُ موردًا لغويًا فرعيًا - على شاكلة مسرد الكلمات.

٣.٢ - المرحلة ٢: إثراء البيانات.

وتعتمد هذه المرحلة على الموارد اللغوية الثانوية للشبكة المنشودة، وتمثّل في:

٣.٢.١ - مُعجم الألفاظ Dictionary (Words Lexicon).

ويُمثّل مصدر المعاني المُعجمية - الحرفية والمجازية - في شبكة الكلمات المنشودة. وتقترح الدراسة جمع مادّته من المُعجمات اللغوية التالية:

- **مُعْجَمُ الْعَرَبِيَّةِ الْحَدِيثَةِ** Arabisches Wörterbuch für die Schriftsprache der Gegenwart للمُستشرق الألماني هانز فير (Hans Wehr) (1981-1909).
- **المُعْجَمُ الوسيط**، الَّذِي أَخْرَجَهُ مَجْمَعُ اللُّغَةِ الْعَرَبِيَّةِ بِالْقَاهِرَةِ فِي طَبْعَتِهِ الْأُولَى عام ١٩٦٠ م، وشارك في تَحْرِيرِهِ نُخْبَةٌ مِنَ الْمُعْجَمِيِّينَ.
- **مُعْجَمُ اللُّغَةِ الْعَرَبِيَّةِ الْمُعاصرة**، الَّذِي حَرَّرَهُ الدُّكْتُورُ أَحْمَدُ مَخْتار (١٩٣٣-٢٠٠٣) بِمُساعدَةِ فَرِيقِ عَمَلٍ، وَظَهَرَتِ الطَّبْعَةُ الْأُولَى مِنْهُ فِي ٢٠٠٨ م.

٢.٢.٣ - مُعْجَمُ التَّعْبِيرَاتِ الاصطلاحية **Idioms Lexicon**.

وهو مصدر المعاني المجازية للتعبيرات الاصطلاحية في الشبكة المنشودة. وتُقدِّمُ الدِّرَاسَةُ جَمْعَ مادَّتهِ مِنَ الْمُعْجَمِيِّينَ التَّالِي ذِكْرُهُمَا:

- **قاموس المتعلم للتعبيرات الكلاسيكية العربية** Learner's Dictionary of Classical Arabic Idioms، الَّذِي حَرَّرَهُ الأَسْتاذُ ليزلي ماكلوكلين (Leslie J. McLoughlin)، ونُشِرَ فِي عام ١٩٨٨ م.
- **المُعْجَمُ السِّيَاقِيّ لِلتَّعْبِيرَاتِ الاصطلاحية**، الَّذِي حَرَّرَهُ الدُّكْتُورُ مُحَمَّدُ إِسْماعيلِ صيني وآخرون، ونُشِرَ فِي ١٩٩٦ م.

٣.٣ - المرحلة ٣: الرِّبْطُ بَيْنَ البِيانَاتِ.

وتُعْنَى هَذِهِ المَرْحَلَةُ بِنِشاءِ الشَّبَكَةِ العِلائِقِيَّةِ Relational Network بَيْنَ مِداخِلِ الشَّبَكَةِ المِنْشُودَةِ. وَيَجْرِي إِنْشاؤها فِي مَرْحَلَتَيْنِ فَرَعِيَّتَيْنِ، نَعْرُضُ لَهَا فِيما يَلِي.

١.٣.٣ - تعيين الحُقُولِ الدَّلَالِيَّةِ.

الحقلُ الدَّلَالِيُّ Semantic field هو مجموعة من الكلمات التي ترتبط دلالاتها وتوضع تحت لفظٍ عامٍّ يجمعها (١). وهناك العديد من التَّصنيفات التي قُدِّمَتْ لتكونَ حُقُولاً دَلَالِيَّةً لمُفْرَدَاتِ اللُّغَةِ؛ لكنَّ أشْمَلَ التَّصنيفاتِ المطروحة وأكثرها منطقيَّةً وملاءمةً لشبكة الكلمات العربيَّة المنشودة - من وجهة نظر الباحث - هو التَّصنيف الَّذِي اقترحه المعجم اليوناني لمُفْرَدَاتِ العهد الجديد Greek New Testament، الَّذِي حرَّره فريقٌ من اللُّغَوِيِّين، حيثُ يَضُمُّ ما يقرب من ٥٠٠٠ كلمة، تحوي قوائمها ١٥٠٠٠ معنى مُعْجَمِيٍّ؛ وتندرج فيه المُفْرَدَاتِ في أربعة حُقُولٍ دَلَالِيَّةٍ (٢)، هي: الموجودات Entities، والأحداث Events، والمُجَرَّدَاتِ Abstracts، والعلاقات Relations. وتتفرَّعُ هذه الحُقُولِ الرَّئِيسِيَّةِ لتشمل حُقُولاً فرعيَّةً، تتفرَّعُ بدورها إلى حُقُولٍ فرعيَّةٍ أُخرى، وهكذا.

وتقترح الدِّراسةُ - إضافةً إلى هذا المعجم - الاستعانة بالمصادر التَّالِيَّةِ:

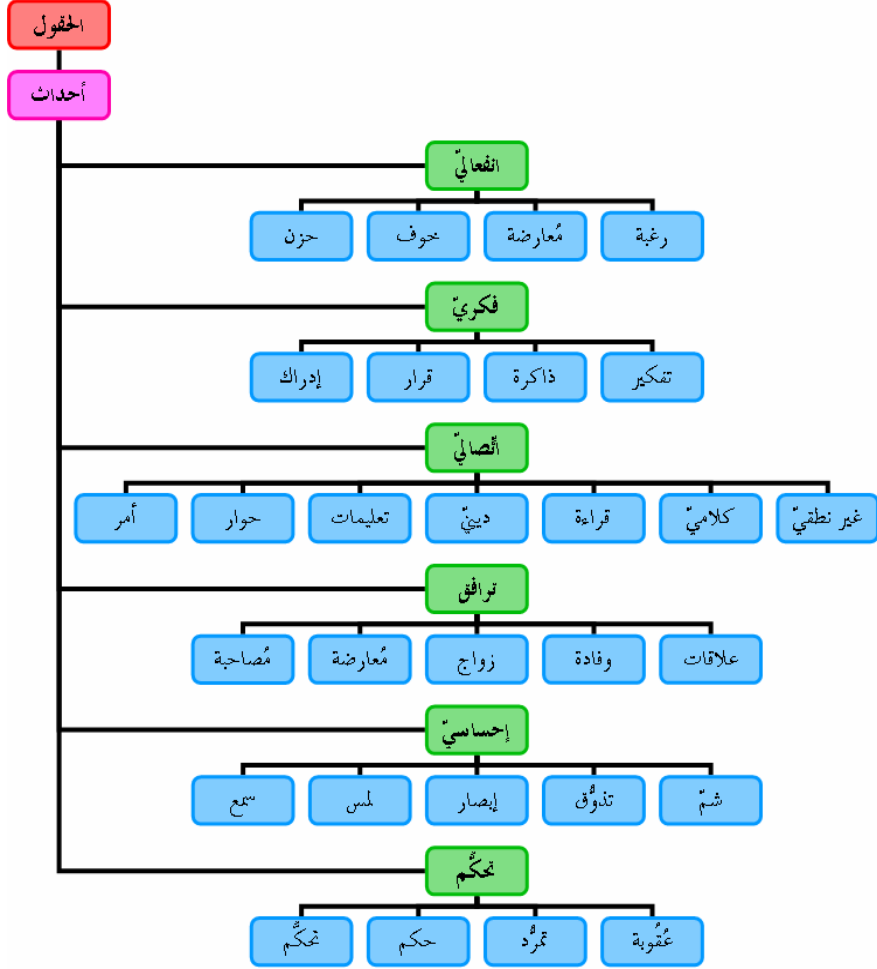
- كتاب «مفهوم النِّظام بوصفه أساساً لصناعة المعجم: محاولة لتصنيف النِّظام» Begriffssystem als Grundlage für die Lexikographie: Versuch eines Ordnungsschemas، الَّذِي حرَّره الألمانِيَّان رودولف هالينج Rudolf

(١٧) () :

(18) United Bible Societies: The Greek New Testament. ISBN 3-438-05110-9.

Hallig وفالتر فون فارتبورج Walther von Wartburg، ونُشِرَ في ١٩٦٣ م.

- مُعجم «المُخَصَّص» لابن سيده الأندلسي (ت ٤٥٨ هـ).
- مُعجم «المنجّد في اللّغة»، لكُراع النمل الهنائي (ت ٣١٠ هـ).

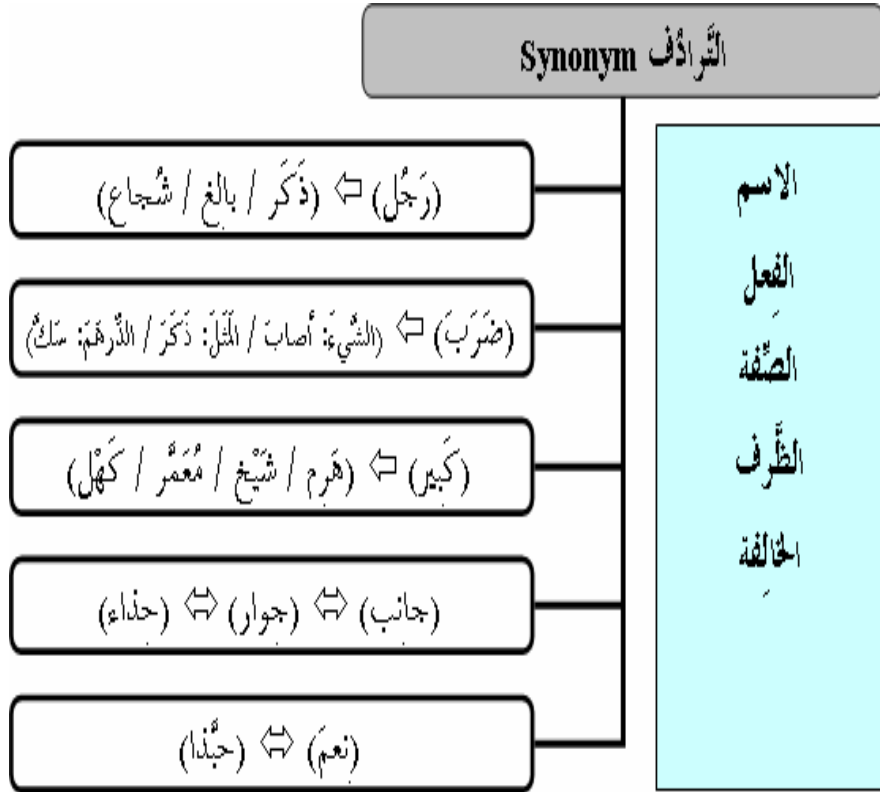


الشكل ٥: نموذج من الحقل الدلالي (الأحداث) - عن مُعجم Greek New

Testament

٢.٣.٣- تعيين العلاقات الدلالية.

تقترح الدراسة الإفادة من العلاقات الدلالية المدرجة في شبكة كلمات برينستون PWN بتطبيقها على مفردات العربية - وفق التقسيم الخماسي، على النحو المبين في المثال التالي (مثال من العربية لعلاقة الترادف في أقسام الكلام الخمسة).



الشكل ٦: علاقة الترادف في أقسام الكلام العربي

وثمة العديد من المصادر التي تُعنى بالعلاقات الدلالية بين مفردات اللغة العربية، والتي يُمكن الإفادة منها في تعيين العلاقات الدلالية بين مفردات شبكة

الكلمات المنشودة؛ وتقرِّحُ الدِّراسةُ مصدرين رَئِيسيين - إضافةً إلى شبكة

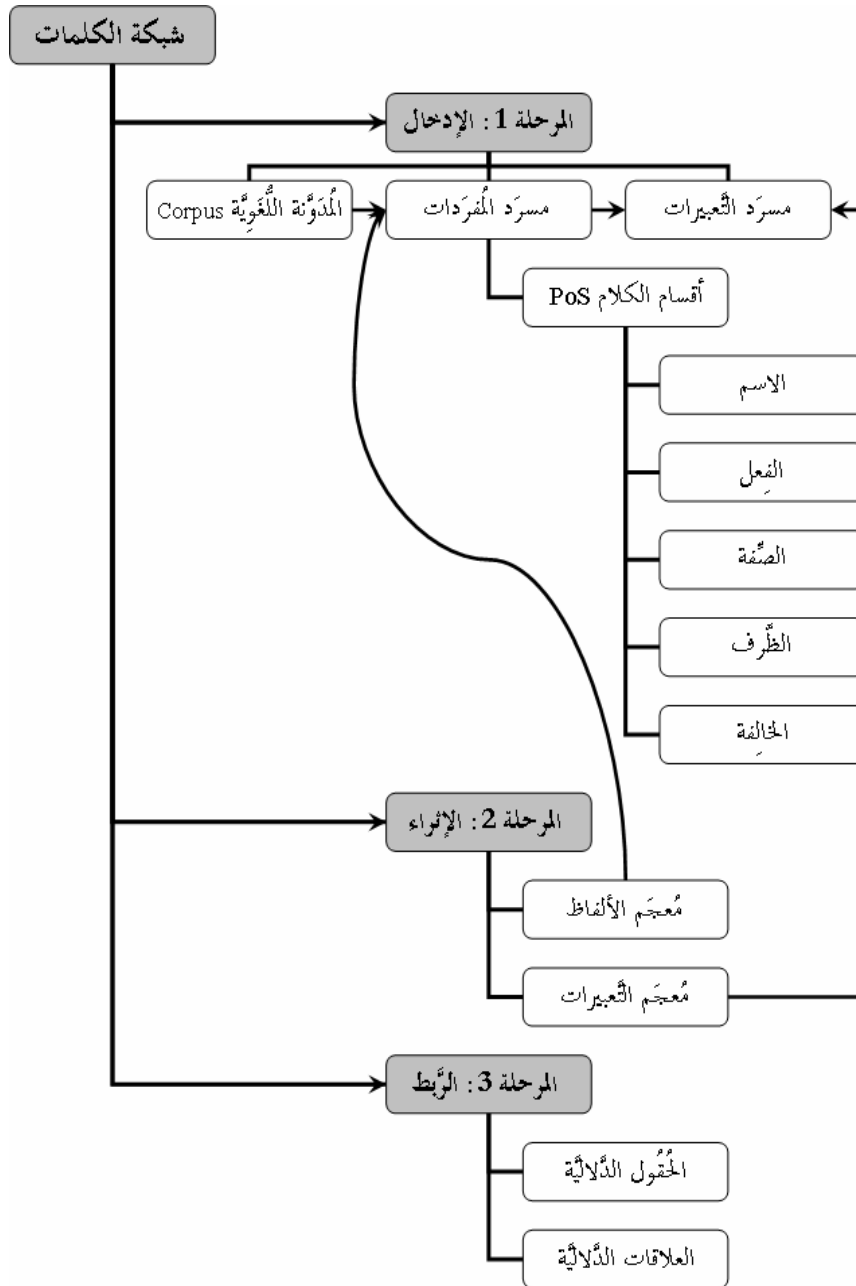
برينستون PWN، هما:

• المكتز الكبير، الَّذي حرَّره الدكتور أحمد مختار عمَر بمُساعدة فريق عمل،

وُنشِرَ في عام ٢٠٠٠ م.

• المكتز العربيِّ المعاصر، الَّذي حرَّره الدكتور محمود إسماعيل صيني

وآخرون، وُنشِرَ في ١٩٩٣ م.



الشكل ٧: الخطوات المنهجية لبناء شبكة الكلمات العربية المنشودة

٤ - التطبيق.

ثمّة العديد من التطبيقات التي تُمثّل جوانب إفادة من شبكة الكلمات العربية المنشودة، والتي ترتبطُ بالية التحليل الدلالي للعربية Semantic analysis. من هذه التطبيقات: تنمية المحتوى الإلكتروني للعربية على الشبكة العنكبوتية Internet - سواءً أكانَ محتوى حُرّاً On-Line أم مُحرَّزاً Off-Line، وتطوير عمل مُحركات البحث العربية Arabic Search engines، والمساهمة في بناء وإثراء الأنطولوجيا العربية المنشودة Arabic Ontology، وتنمية بيئة الويب الدلالي، وتطبيقات صناعة المعجم العربي؛ والتطبيق الأخير هو الذي يعيننا في هذه الدراسة.

إنّ الاعتمادَ على الموارد اللغوية (المُدونة اللغوية، ومُعجم المُفردات / الألفاظ، ومُعجم التعبيرات الاصطلاحية) يُثمرُ هيكلًا مُعجميًا يتوافق مع مُعجمات الألفاظ؛ كما يُثمرُ الربطُ بين الموارد اللغوية باستعمال الحُقُول الدلالية - من ناحية - والعلاقات الدلالية - من ناحية أخرى - هيكلًا مُعجميًا يتوافق مع مُعجمات الموضوعات.

ويمكنُ القولُ بعبارةٍ أخرى: إنّ الالتزامَ بمعايير بناء شبكات الكلمات - مع مُراعاة مُناسبتها لطبيعة اللغة العربية على مُستوى المباني والمعاني - عندَ إعداد شبكة الكلمات العربية المنشودة سيُثمرُ قاعدة بياناتٍ مُعجمية هائلة تشكّل في هيكلٍ مُعجميٍّ ضخمٍ، يجمعُ بين مُعجمات الألفاظ ومُعجمات الموضوعات،

ويصلح مورداً لغوياً للعديد من التطبيقات المتعلقة بقضايا اللغة العربية عموماً وتقنياتها على وجه الخصوص.

ومن هذا الهيكل المعجمي الذي يتشكل في مجموعة من الحُقُول / المجالات الدلالية، نستطيع بناء العديد من المعجمات اللغوية المتخصصة، التي يدعو إليها تنامي المعارف ونُشوء العديد من العلوم البينية (كحوسبة اللغة، والهندسة الطبية، والرياضيات الفيزيائية)؛ وهذا مما يدعم تعريف العلوم ويساعد على إثراء المحتوى العربي.

٥- نتائج الدراسة.

وبعد... فقد خلصت الدراسة إلى النتائج التالية:

١. رأت الدراسة أن شبكة الكلمات العربية المنجزةAWN ليست شبكة للكلمات بالمفهوم الدقيق، وإنما هي ترجمة أو محاكاة لبعض أجزاء الشبكة الإنجليزية «برينستون» PWN؛ وقد أدت هذه الترجمة إلى قُصور منهجي، تمثل في: اعتبار العربية لغةً لصقية، وإغفال أثر البيئة في اللغة.
٢. اقترحت الدراسة منهجيةً لبناء شبكة للكلمات العربية انطلاقاً من شبكة كلمات برينستون PWN، مع مراعاة طبيعة اللغة العربية. وتقوم المنهجية المقترحة على ثلاث مراحل أساسية، هي (الإدخال، والإثراء، والربط بين البيانات).
٣. اقترحت الدراسة تقسيم الكلام في الشبكة المنشودة إلى خمسة أقسام، هي: الاسم، والفعل، والصفة، والظرف، والخالفة.

٤. اقترحت الدراسةُ الإفادة من مواردٍ لغويّةٍ مُعيّنة في تعبئة شبكة الكلمات المنشودة وإثراء محتواها، وتمثّلت في: (المُدونة اللُّغويّة، ومسرّدي المفردات والتّعبيرات، ومُعجمي المفردات والتّعبيرات، ومُعجمي الحُقول الدّلاليّة والعلاقات الدّلاليّة).
٥. عرّضت الدراسةُ لبعض تطبيقات شبكة الكلمات، لا سيّما في صناعة المُعجم.
- ٦- الخُلاصة.

سعت هذه الورقةُ لِ التّمييز بين المُصطلحات الثّلاثة (شبكة الكلمات، والشّبكة الدّلاليّة، والأنطولوجيا)؛ وعرّضت الدراسةُ لإشكاليّة شبكة الكلمات العربيّة، كما قدّمت منهجًا مُقترَحًا لبناء وإثراء الشّبكة المنشودة بما يتوافق مع طبيعة اللّغة العربيّة، وبما يتناسب مع تطبيقات الشّبكة في تقنيات اللّغة العربيّة وصناعة المُعجم العربيّ.

٧- المراجع.

- (1) Fellbaum, C., (1998). WordNet: an electronic lexical database. Cambridge, Mass: MIT Press
- (2) Black, W., Elkateb, S., Rodriguez, H, Alkhalifa, M., Vossen, P., Pease, A. and Fellbaum, C., (2006). Introducing the Arabic WordNet Project, in Proceedings of the Third International WordNet Conference, Sojka, Choi, Fellbaum and Vossen eds.
- (3) Staab, S., Studer, R., (2004). Handbook on ontologies. Birkhäuser.
- (4) Poli, R., Healy, M., Kameas, A., (2010). Theory and Applications of Ontology: Computer Applications. Springer.
- (5) Finegan, E., (2011). Language: its structure and use. Thomson Wadsworth. 6th Edition.
- (6) Ginsburgh, V., Weber, S., (2011). How many languages do we

need?: the economics of linguistic diversity. Princeton University

Press.

(٧) السَّعِيد (المُعْتَزَّ بالله): اللُّغَةُ الْإِنْسَانِيَّةُ، دار الهاني للنَّشر، القاهرة، ط ١، ٢٠١١ م.

(8) <http://www.illc.uva.nl/EuroWordNet/>

(9) <https://gforge.inria.fr/projects/wolf/>

(10) <http://www.ontologyportal.org/>

(11) <http://www.mpi-inf.mpg.de/yago-naga/yago/>

(١٢) السَّعِيد (المُعْتَزَّ بالله): مُدَوَّنَةٌ مُعْجَمٌ تَارِيخِيٌّ لِلُّغَةِ الْعَرَبِيَّةِ «مُعَالَجَةٌ لُغَوِيَّةٌ حَاسُوبِيَّةٌ»،

رسالة دكتوراه، جامعة القاهرة، ٢٠١١ م.

(١٣) حَسَّان (تَمَّام): مَنَاهِجُ الْبَحْثِ فِي اللُّغَةِ، مَكْتَبَةُ الْأَنْجَلُو، الْقَاهِرَةَ، ط ١، ١٩٥٥ م.

(١٤) أَنيس (إِبْرَاهِيم): مِنْ أَسْرَارِ اللُّغَةِ، مَكْتَبَةُ الْأَنْجَلُو الْمِصْرِيَّةِ، الْقَاهِرَةَ، ط ٤، ١٩٧٢ م.

(١٥) حَسَّان (تَمَّام): اللُّغَةُ الْعَرَبِيَّةُ «مَعْنَاهَا وَمَبْنَاهَا»، الْهَيْئَةُ الْمِصْرِيَّةُ الْعَامَّةُ لِلْكِتَابِ، الْقَاهِرَةَ،

ط ٢، ١٩٧٩ م.

(١٦) عُمَرُ (أَحْمَدُ مُحْتَار): عِلْمُ الدَّلَالَةِ، عَالَمُ الْكُتُبِ، الْقَاهِرَةَ، ط ٥، ١٩٩٨ م.

(17) United Bible Societies: The Greek New Testament. ISBN 3-438-05110-9.