نحوَ شبكةٍ للكلمات العربيَّة لأغراض الصِّناعة المُعجميَّة

المُعتزّ بالله السَّعيد
كُلِّيَّة دار العُلُوم، جامعة القاهرة، مصر
moataz@cu.edu.eg

المستخلص - تَسعى هذه الدِّراسةُ إلى تقديم منهجيَّةٍ لبناء وتطوير شبكة للكلمات العربيَّة لأغراض الصِّناعة المُعجميَّة.وتهدفُ الدِّراسةُ إلى إيجاد رؤية واضحة لآليَّات بناء شبكةٍ للكلمات العربيَّة، تُراعي طبيعةَ اللُّغة العربيَّة ونظامَها المُعجميّ؛ بحيثُ يُمكنُ توظيفُ هذه الآليَّات مُستقبلًا في مختلف أشكال الصِّناعة المُعجميَّة العربيَّة. وتأتي الدِّراسةُ في ستَّة محاورَ تتضمَّن مُقدِّمةً ثُمَّ عرضًا لإشكالات الدِّراسة؛ ويلي ذلك تقديمُ المنهجيَّة المُقترَحة لبناء شبكة الكلمات العربيَّة، ثُمَّ عرضٌ لكيفيَّة توظيف شبكة الكلمات في صناعة المُعجم العربيّ. وأخيرًا يستعرضُ الباحثُ نتائجَ الدِّراسة، ويعرضُ خُلاصةَ بحثه.

الكلمات المفاتيح
شبكة الكلمات – الأنطولوجيا – صناعة المُعجم – العلاقات الدِّلاليَّة

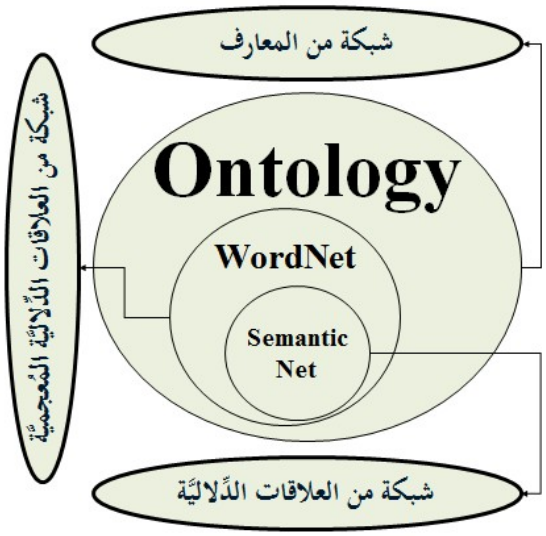
1 المقدمة

1.1 في ماهيَّة شبكة الكلمات

يُمكنُ تعريفُ شبكة الكلمات WordNetبأنَّها قاعدةُ بياناتٍ مُعجميَّة دلاليَّة Lexical-Semantic Database، تُعالَجُ فيها مُفرَداتُ اللُّغة بتعيين أقسامِ الكلام Parts of Speech (PoS) وفقَ ما يقتضيه النِّظامُ المُعجميُّ للُّغة المُعَيَّنة، ثُمَّ تُصَنَّفُ في مجموعاتٍ مُترابطةٍ من المترادِفات Synonyms (Synsets) الَّتي تُعَبِّرُ عن مَفاهِيمَ ثابتة، وترتبطُ فيما بينها بشبكةٍ من العلاقات الدِّلاليَّة Semantic Relations. وتُعبِّرُ شبكةُ الكلمات عن مُجتمعٍ لُغويٍّ مُتجانسٍ من المُفردات والمعاني، حيثُ تتَّفقُ فيها أنماطُ الكلمات ومعانيها الوظيفيَّة والمُعجميَّة، وإن اختلَفَت على مُستوى السِّياق اللُّغويّ [3، 4].

وقد ظَهَرَت فِكرَةُ شَبَكة الكلماتفي اللُّغَة الإنجليزِيَّة أوَّلًا في الثَّمانينيَّات من القرن العشرين،ثُمَّ انتقلَت منها إلى مُختلف اللُّغات الإنسانيَّة، بما فيها العربيَّة؛ لتكونَ موردًا وأداةً للبحث والتَّطوير في عُلُوم اللُّغة والمُعجم وميادين المعلوماتيَّة. وأمكنَ توظيفُ شبكات الكلمات في الصِّناعة المُعجميَّة الحديثة، بما تحويه من ثروةٍ لُغَويَّةٍ من المُفرَدات والمعاني والعلاقات الدِّلاليَّة، كما أمكَنَ توظيفُها في تطبيقات الذَّكاء الاصطناعيّ Artificial Intelligence (AI) ومُعالَجَة اللُّغات الطَّبِيعِيَّة Natural Language Processing (NLP) [3، 16].

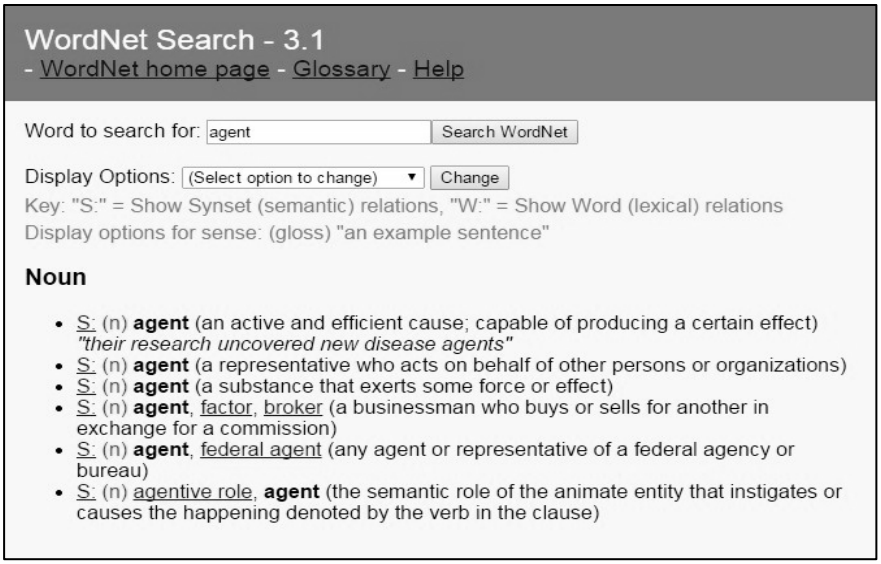
وتتكوَّن شبكةُ الكلمات من مجموعةٍ من الشَّبكات الدِّلاليَّة Semantic Nets الَّتي تربطُ بينَ المفاهيم باعتبارِها شكلاً من أشكال التَّمثيل المَعرِفيّ؛ وفي الوقت ذاتِه ينتُجُ هيكلٌ مَعرِفِيٌّ ضخم عن الارتباط بين مجموعةٍ من شبكات الكلمات، يُستَخدَمُ في تطبيقاتٍ عديدةٍ تتبعُ ميادينَ الذَّكاءِالاصطناعيِّ وعلومِ المكتَبات وتطبيقات الويب الدِّلاليّ Semantic web، ويُطلَقُ عليه "الأنطولوجيا Ontology" [11]. وتُمَثِّلُ الأنطولوجيا مجموعةً من الكُتَل المَعرِفِيَّة الَّتي تتلاحمُ عناصِرُها لتُكَوِّنَ تمثيلاً لمجموعاتٍ مُنَظَّمة من المفاهيم المُشتركة في حُقُول دلاليَّة مُحدَّدة، مُترابطة بعلاقات دِلاليَّةٍ، وتُستَقَى مادَّتُها من مُختلف فُرُوع المعرِفة شريطَةَ أن تَعكِسَ واقِعَ اللُّغة المُعَيَّنة [4].ومع وُجود فرقٍ بينَ شبكة الكلمات والشَّبكة الدِّلاليَّة من ناحيةٍ، وبينَ شبكة الكلمات والأنطولوجيا من ناحيةٍ أخرى، إلَّا أنَّ بعضَ المشرُوعات لا تُميِّزُ بينها، الأمرُ الَّذي أدَّى إلى تداخُلٍ بينَ مفاهيم المُصطلحات الثَّلاثة.



الشكل رقم (1)
العلاقة بينَ شبكة الكلمات WordNet والشَّبكة الدّلاليّة Semantic Net والأنطولوجيا Ontology

1.2 الخلفيَّة والإرهاصات

ظهرَت فكرةُ شبكة الكلمات WordNet في عام 1985م، حيثُ بدأَ العملُ في بناء شبكةٍ لكلمات اللُّغة الإنجليزيَّة في مُختَبَر العُلُوم الإدراكِيَّة بجامعة برينستون Princeton الأمريكيَّة، بهدف توفير موردٍ مُعجميٍّ يربطُ بينَ معاني الكلمات في الإنجليزيَّة. وتَمَكَّنَ الفَريقُ الَّذي أشرَفَ عَلَيه عالم النَّفس الأمريكِيّ جورج ميلر George Miller(1920-2012) مِن بِناء الشَّبكة الموسُومة بـ "شبكة كلمات برينستون Princeton WordNet" في غُضُون عشرين عامًا، بدعمٍ من مؤسَّساتٍ حُكوميَّة أمريكيَّة، على رأسِها: المُؤسَّسة الوطنيَّة للعُلُوم NSF ووكالة مشرُوعات البُحُوث المُتطوِّرة الدِّفاعيَّة DARPA. وقامَ فريقُ العمل في شبكة كلمات برينستونPWN بحوسَبة مادَّتِها لتتوافقَ مع أنظمة تشغيل الحواسيب، قبلَ إتاحتِها على الشَّبَكة العنكبوتيَّة في عدَّة إصداراتٍ، آخرُها الأصدار 3.1 في عام 2012م. واشتَمَلَت قاعدةُ بيانات هذا الإصدار على 155287 كلمة مُفرَدة تَضُمُّها 117659 مجموعة ترادُفيَّة Synset [12].



الشكل رقم (2)
نموذج من شَبَكَة كلمات برينستون لِلُّغة الإنجليزِيَّة – الإصدار 3.1

وقد مهَّدَت شبكةُ كلمات برينستونPWN الطَّريقَ أمامَ بناء العديد من الشَّبكات الأخرى للُّغات الإنسانيَّة، حيثُ اتُّخِذَ منهجُها معيارًا لرابطة شبكات الكلمات العالميَّة Global WordNet Association (GWA) الَّتي تهدفُ إلى إحداث التَّجانُس بينَ شبكات الكلمات وتوحيد أشكالها في مُختلف اللُّغات. وتحتَ مظلَّة هذه الرَّابطة، أمكنَ إنجازُ شبكاتٍ أخرى، مثل: شبكة الكلمات الأوروبيَّة Euro WordNet (EWN)، وشبكة الكلمات الحُرَّة للفرنسيَّة WordNet Libre du Français (WOLF)، وشبكة الكلمات الفنلنديَّة FinnWordNet، وشبكة الكلمات الصِّينيَّة Chinese Open WordNet،وشبكة الكلمات العبريَّة Hebrew WordNet، وشبكة الكلمات مُتعدِّدة اللُّغات MultiWordNetوالأنطولوجيا العُليا المُوَحَّدة - سومو Suggested Upper Merged Ontology (SUMO)، وأنطولوجيا ياجو Yet Another Great Ontology (YAGO)، وغيرِها [3، 4].

1.3 شبكة الكلمات في اللُّغة العربيَّة

عَرَفَت اللُّغةُ العربيَّةُ طريقَها إلى شبكة الكلمات في مطلع الألفيَّة الثَّالثة. وظهرَت عدَّةُ مشروعاتٍ لبناء وتطوير شبكاتٍ للكلمات العربيَّة، منها: مشروعُ "شبكة الكلمات العربيَّة Arabic WordNet (AWN)" الَّذي قامَ به فريقٌ بحثيٌّ في جامعات برينستون وبرشلونة ومانشستر بينَ عامَي 2005 و 2007م، بإشراف الباحثةِ الألمانيَّة كريستيان فيلباوم Christiane Fellbaum– الَّتي شارَكَت من قبلُ في بناء شَبَكَة كلِمات برينستون PWN تحتَ قِيادَة جورج ميلر. وتقومُ فكرةُ شبكة الكلمات العربيَّة AWN على بناء قاعدة بياناتٍ علائقيّة Relational Database من المُفرَدات المُترابطة فيما بينها بعلاقاتٍ دلاليَّةٍ في صُورة بناءٍ شجريٍّ تُصَنَّفُ المُفرَداتُ فيه إلى قِسمَي الكلام (الاسم، والفِعل)، ويتفرَّعُ عن كلِّ قسمٍ مجموعةٌ من الحُقُول الدِّلاليَّة Semantic Fields الَّتي تضُمُّ عددًا من المُفرَدات المُوَزَّعة على حُقُولٍ فرعيَّةٍ. واشتملَت هذه الشَّبكةُ على 17561 كلمة مُفرَدة تَضُمُّها 7822 مجموعة ترادُفيَّة Synset [10]. ومع ضعف المُحتوى المُتَضَمَّن في شبكة الكلمات العربيَّة AWN مُقارنةً بغيرها من الشَّبكات المُنجَزة للُّغاتٍ أخرى، فقد سعى فريقُ العمل إلى تطويرها وإثراء قواعد بياناتِها في مُحاولةٍ للوُصُول بها إلى صُورةٍ تسمحُ بالإفادة الفعليَّة منها، اعتمادًا على خوارزميَّاتٍ وموارِدَ لُغويَّةٍ للعربيَّة المُعاصرة. ومن المُزمَع أن تصدُرَ شبكةُ الكلمات العربيَّة AWN في نُسخةٍ مُحسَّنةٍ عام 2015 [9].

وفي الآونة الأخيرة ظَهَرَت بعضُ المشرُوعات الأخرى الَّتي تُحاولُ استدراكَ القُصُور في شبكة الكلمات العربيَّة، منها: مشروع "أنطولوجيا اللُّغة العربيَّة Arabic Ontology (AO)" الَّذي ترعاهُ جامعة بيرزيت في فلسطين منذ عام 2011، ويُشرفُ عليه الأكاديميّ الفلسطينيّ مُصطفى جرَّار؛ ويسعى القائمُونَ على المشرُوع إلى بناء شبكةٍ من المفاهيم العربيَّة الَّتي ترتبطُ بعلاقاتٍ دلاليَّةٍ في بناءٍ شجريٍّ يجمعُ معاني الكلمات العربيَّة. وعليه، فالمشرُوعُ لا يهدفُ إلى بناء شبكةٍ للكلمات العربيَّة، وإنَّما يهدفُ إلى بناء شبكةٍ من معاني الكلمات. وتنطلقُ "أنطولوجيا اللُّغة العربيَّة" من التَّجارب السَّابقة لبناء شبكات وأنطولوجيات اللُّغات الإنسانيَّة، لاسيَّما "شبكة كلمات برينستون PWN" و "الأنطولوجيا العُليا المُوَحَّدة SUMO" مع التَّركيز على العلاقات المنطقيَّة بينَ معاني الكلمات بصرف النَّظَر عن طبيعة اللُّغة العربيَّة الاشتقاقيَّة.

ومن ناحيةٍ أخرى، أعلَنَت المُنظَّمة العربيَّة للتَّربية والثَّقافة والعُلُوم ALECSO في مطلع عام 2013 عن مشروع "الشَّبكة الدِّلاليَّة للُّغة العربيَّة SEmantic Wordnet of ArabiC (SEWAC)" برعايتِها وإشراف الباحث المغربيّ كريم بُوزُوبع، بهدف توفير موردٍ لُغويٍّ مفتوح المصدر يُستفادُ منهُ في مُختلَف أغراض البحث العلميّ. ويعتمدُ هذا المشرُوعُ في تعيين مفاهيم الكلمات على "المُعجم الحاسُوبيّ التَّفاعُليّ" الَّذي أنجَزَهُ فريقٌ من الباحثينَ برعاية مدينةالملكعبدالعزيزللعُلُوموالتِّقنيةبالرِّياض KACSTوالمُنظَّمة العربيَّة للتَّربية والثَّقافة والعُلُوم ALECSO، كما يعتمدُ في تعيين العلاقات الدِّلاليَّة على "المكنز الكبير" الَّذي صنعَهُ فريقٌ من الباحثينَ بإشراف الدُّكتور أحمد مُختار عُمَر. ومع أنَّ

المشروعَ لايزالُ قائمًا منذ ما يربو على عامَين، إلَّا أنَّهُ لم يخطُ خُطواتٍ تُذكر إلى الآن، باستثناء إعداد منهجيَّتِه وتوفير بعض مواردِه وأدواته. ويعودُ هذا إلى ضعف التَّمويل المُتاح للمشروع، بالإضافة إلى اعتماده أساسًا على جُهود أساتذةٍ وباحثينَ مُتطوِّعينَ غير مُتفرِّغين.

مَثَّلَت هذه المشرُوعاتُ مُحاولاتٍ وتجاربَ لمؤسسَّاتٍ علميَّةٍ لا تهدفُ إلى الرِّبح. وعليه، فالمُفتَرَضُ أن تُتاحَ مُخرَجاتُها في صُورةٍ حُرَّةٍ أو مفتوحة. لكنَّ ثمَّةَ تجاربَ أخرى، أنجزَتها شركاتٌ مَعنيَّةٌ بحوسبة اللُّغة العربيَّة وتنمية مواردِها. ونستطيعُ التَّمثيلَ لهذه التَّجارب بـ "قاعدة بيانات صخر المُعجميَّة الدِّلاليَّة" الَّتي أنجَزَتها شركة "صخر" في العقد الأوَّل من الألفيَّة الثَّالثة لاستخدامِها في تطوير أدواتها؛ وهي مُستمدَّةٌ من مُدَوَّنةٍ لُغويَّةٍ للعربيَّة المُعاصرة، ومن معاجمَ عربيَّة مُتنوِّعة [15]. وكذلكَ "قاعدة بيانات الرَّبط الدِّلاليّ المُعجميّ في اللُّغة العربيَّة" الَّتي أنجَزَتها الشَّركة الهندسيَّة لتطوير النُّظُم الرَّقميَّة RDI بينَ عامَي 2005 و 2011؛ واشتمَلَت هذه القاعدةُ على نحو مئة ألف مُفردة ترتبطُ فيما بينها باثنتين وعشرينَ علاقة دلاليَّة، وتُفيدُ منها الشَّركةُ كذلكَ في تطوير أدواتها وموارِدِها اللُّغَويَّة.

2 إشكالات بناء شبكة للكلمات العربيَّة

إذا أردنا أن نضَعَ منهجًا واضحًا لبناء شبكةٍ للكلمات العربيَّة يُستفادُ منها في أغراض الصِّناعة المُعجميَّة، فعلينا أوَّلًا أن نقفَ على إشكالات بناء هذه الشَّبكة، الَّتي نستطيعُ استخلاصَها من المُحاولات السَّابقة لبناء شبكاتٍ للكلمات العربيَّة في ضوء طبيعة اللُّغة. لقد عرضنا آنفًا لثلاث مُحاولاتٍ، تمثَّلَت في: شبكة الكلمات العربيَّة AWN والأنطولوجيا العربيَّة AO والشَّبكة الدِّلاليَّة للُّغة العربيَّة SEWAC؛ وهي المُحاولاتُ الَّتي تستدعي الوُقُوفَ عليها لتقييمها. أمَّا تجاربُ الشَّركات فلا يعنينا أن نقفَ عليها لأمرين، الأوَّلُ أنَّها مُغلقةٌ ولا يُسمَحُ بالإفادة منها إلَّا فيما يخدمُ أغراضَ الجهة الَّتي صَنَعَتها. والأمرُ الآخَرُ أنَّها ليسَت شبكاتٍ للكلمات بالمفهُوم الدَّقيق؛ وإنَّما هي موارد لُغَويَّة ومُعجميَّة وُضِعَت في صُورة أنطولوجياتٍ جُزئيَّة. وبالنَّظر إلى المُحاولات المُنجَزة سلفًا مع مُراعاة طبيعة اللُّغة العربيَّة، نستطيعُ أن نقفَ على ثلاثة إشكالاتٍ رئيسةٍ، نعرضُها فيما يلي.

1. طبيعة النِّظام المُعجميّ للعربيَّة

يقتضي النِّظامُ المُعجميُّ للُّغة العربيَّة أن نُمَيِّزَ بينَ المداخل المُعجميَّة Lexical Entries والوحدات Lexemes الَّتي تنسدلُ عنها، ويقتضي هذا النِّظامُ – كذلك – أن نُقَسِّمَ الوحداتِ المُعجميَّةَ باعتبارِها أقسامَ الكلام إلى: الاسم والفعل والأداة، بحيثُ تختصُّ الأسماءُ والأفعالُ بالمعاني المُعجميَّة وتختصُّ الأدواتُ بالمعاني الوظيفيَّة. وبالنَّظر إلى المُحاولات الثَّلاثة المذكورة آنفًا، سنجدُ أنَّ شبكة الكلمات العربيَّة AWN قد سارَت في مادَّتِها على منهج شبكة كلمات بريستون PWN، فظهرت كقطعةٍ مُعرَّبةٍ منها، تتعاملُ مع العربيَّة باعتبارها لُغةً التصاقيَّة، الأمرُ الَّذي يتنافى مع النِّظام المُعجميّ للعربيَّة؛ ذلك أنَّها لا تُميِّزُ بينَ المدخل المُعجميّ من ناحية، وتكتفي بتقسيم الكلام إلى الاسم والفعل من ناحيةٍ أخرى. ولا يختلفُ الحالُ كثيرًا في الأنطولوجيا العربيَّة AO؛ إذ اعتمدَت على أسُسٍ منطقيَّةٍ بصرف النَّظر عن المُعطيات اللُّغَويَّة. أمَّا الشَّبكة الدِّلاليَّة للُّغة العربيَّة SEWAC فاكتَفَت في مادَّتها بالكلمات الَّتي لها علاقةٌ دلاليَّةٌ مع غيرها، دونَ التَّمييز بينَ المداخل والوحدات.

2. طبيعة اللُّغة العربيَّة في مبانيها

تتشكَّلُ ألفاظُ اللُّغة العربيَّة في أربعةٍ وثلاثين فونيمًا تجمعُ بينَ الصَّوامت والصَّوائت، وينتج عن التَّداخل بين الفونيمات (المنطوقة) تداخُلٌ بين الجرافيمات (المكتوبة)؛ وهذا ما يؤدِّي إلى تغيُّر شكل الحرف الواحد في الكلمة تبعًا لموضعه فيها، ويؤدِّي كذلك إلى تعدُّد أشكال الكلمة المُتَّفقةِ في حُرُوفها. والواقعُ أنَّ المُحاولاتِ السَّابقة لم تهتمَّ بأشكال الكلمة المضبوطة، فلم نجد – مثلًا – تمييزًا بينَ أشكال الكلمة (بل) الَّتي يُمكنُ أن تردَ اسمًا أو فعلًا أو أداة، فتأتي على الصُّورة (بَلٌّ) للدِّلالة على صيغة المصدر،

وعلى الصُّورة (بَلَّ) للدِّلالة على صيغة الفعل، وعلى الصُّورة (بَلْ) للدِّلالة على حرف العطف. وعليه، أهملت المُحاولاتُ السَّابقةُ القدرَ الأكبرَ من الكلمات العربيَّة، ما أدَّى إلى نقصٍ كبير في مُحتواها.

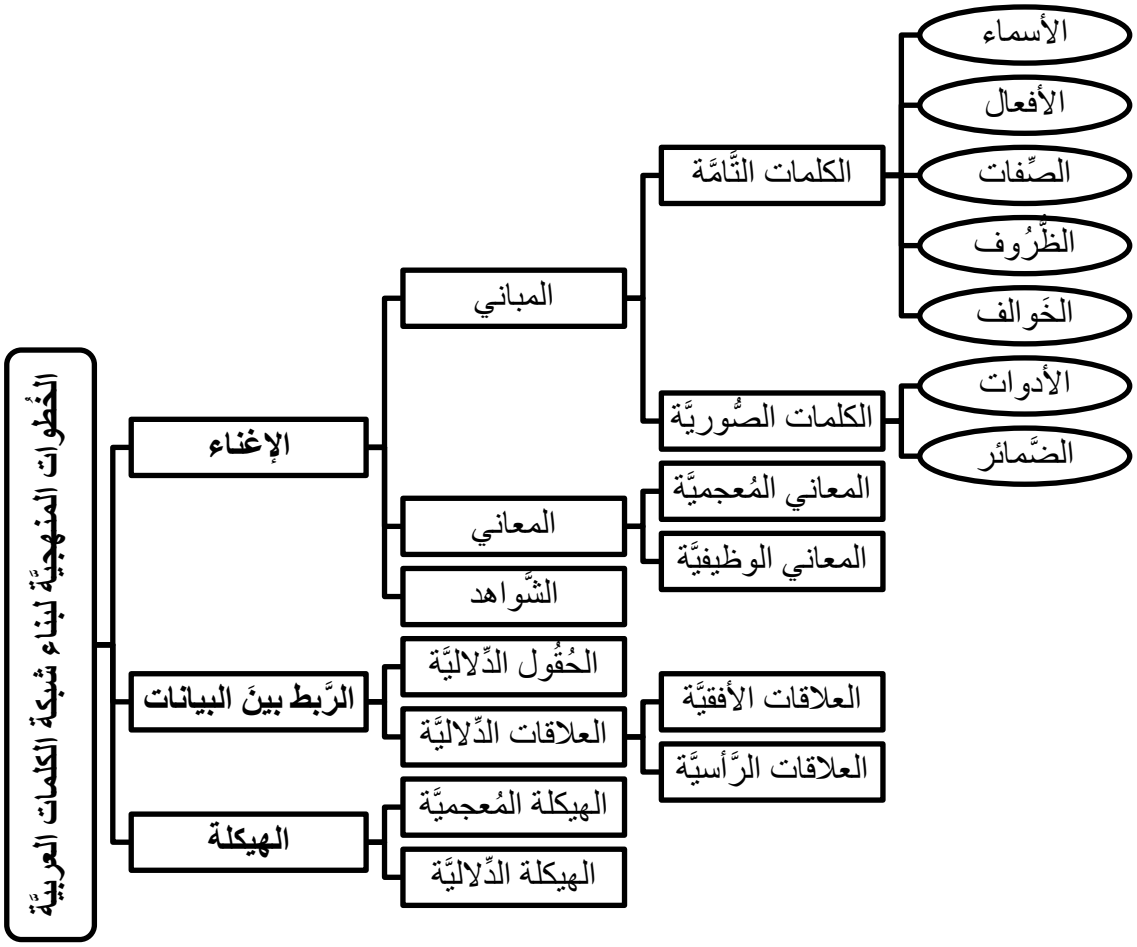
3. طبيعة اللُّغة العربيَّة في معانيها

يبدو أثرُ البيئة العربيَّة واضِحًا في تعدُّد المُفرَدات الَّتي تُشِيرُ إلى معنى واحد. ونستطيعُ التَّمثيلَ على ذلك بالكلمات (سَبع، غَضَنفَر، حارِث، أُسامة،...)؛ إذ تُشِيرُ جميعُها إلى (الأسَد). ووفقًا لمناهج الشَّبكات المُنجَزة، لن يكونَ لهذه الكلمات وجودٌ في قاعدة بياناتها؛ لأنَّها مُستمدَّةٌ من البيئة العربيَّة. فالواقعُ أنَّ المُحاولاتِ السَّابقةَ سَعَت إلى مُحاكاة اللُّغة الإنجليزيَّة في مُعالجة المعاني، فبَدَت مُعبِّرةً عن لُغةٍ أخرى. وإذا استطعنا تمييزَ دلالات الكلمات الَّتي تشتركُ في البيئتين العربيَّة والإنجليزيَّة، فإنَّنا لن نتمكَّنَ من التَّعامُلَ مع كلماتٍ تخصُّ البيئةَ العربيَّةَ دونَ غيرها. وعلى سبيل المثال، سنستطيعُ التَّعبيرَ عن دلالات كلمة (أسَد) الَّتي تجمعُ الكلمات ("أسد" للمُذكَّر، و "لَبُؤَة" للمُؤَنَّث، و "شِبل" للصَّغير) بالمُقابلات الإنجليزيَّة ("Lion" للمُذكَّر، و "Lioness" للمُؤَنَّث، و "Cub" للصَّغير) على التَّرتيب، لكنَّنا لن نتمكَّنَ من التَّعبير عن دلالات كلمة (جَمَل) الَّتي تجمعُ الكلمات الثَّلاثة ("جَمَل" للمُذكَّر، و "ناقة" للمُؤَنَّث، و "حُوار" للصَّغير) إلَّا بمُقابل إنجليزيٍّ واحدٍ، هو"Camel".

3 منهجيَّة بناء شبكة للكلمات العربيَّة

لمَّا كانت الغايةُ من هذه الدِّراسةِ توفيرَ مَوردٍ لُغَويٍّ يُستفادُ منهُ في أغراض الصِّناعة المُعجميَّة العربيَّة، لَزِمَ أن نقفَ على ما تتطلَّبُهُ هذه الصِّناعةُ، استئناسًا بواقِعِها وما تُعانيه من جوانب القُصُور. والمُفتَرَضُ في مُحتوى شبكة الكلمات المنشُودة أن يكونَ مُناسبًا للمنطق اللُّغَويّ ومُوافقًا لطبيعة اللُّغة العربيَّة الاشتقاقيَّة ونظامِها المُعجميّ من ناحية، وثريًّا بما يُلبِّي حاجةَ المُعجمات العربيَّة من ناحيةٍ أخرى. ويستدعي تحقيقُ هذا الأمر أن تكونَ الشَّبكةُ المنشُودةُ تمثيلًا لمراحل اللُّغة العربيَّة (القديمة والوسيطة والمُعاصِرة) ليُستفادَ منها حالَ توظيفِها في صناعة مُعجمٍ تاريخيٍّ أو مُعجماتٍ مرحليَّةٍ للعربيَّة، وأن تتضمَّنَ قدرًا كبيرًا من العلاقات الدِّلاليَّةِ بينَ مُفرَداتِها ليُستفادَ منها حالَ توظيفِها في صناعةِ مكنزٍ لُغويٍّ أو مُعجمٍ موضُوعيّ للمعاني. وفي كُلِّ الأحوال ينبغي أن تضُمَّ قدرًا من الشَّواهد والأمثلة الَّتي تعكسُ الاستعمالَ اللُّغَويَّ للمُفرَدات.

وتحقيقًا لهذه الغاية، تقترحُ الدِّراسةُ منهجيَّةً لبناء شبكة الكلمات المنشودة في ثلاث مراحلَ أساسيَّة، حيثُ تُعنى المرحلةُ الأولى بتوفير موارد شبكة الكلمات المنشودة، وتُعنى المرحلةُ الثَّانيةُ بالرَّبط بينَ الوحدات اللُّغَويَّة [المُفرَدات] المُتَضَمَّنة في شبكة الكلمات باستخدام الحُقُول والعلاقات الدِّلاليَّة، وتُعنى المرحلةُ الثَّالثةُ والأخيرةُ ببناء هيكل شبكة الكلمات وفقَ ما يُحقِّقُ غايةَ الإفادة منها في الصِّناعة المُعجميَّة العربيَّة، وبيانُ ذلك على النَّحو المُبيَّن في (الشَّكل 3):



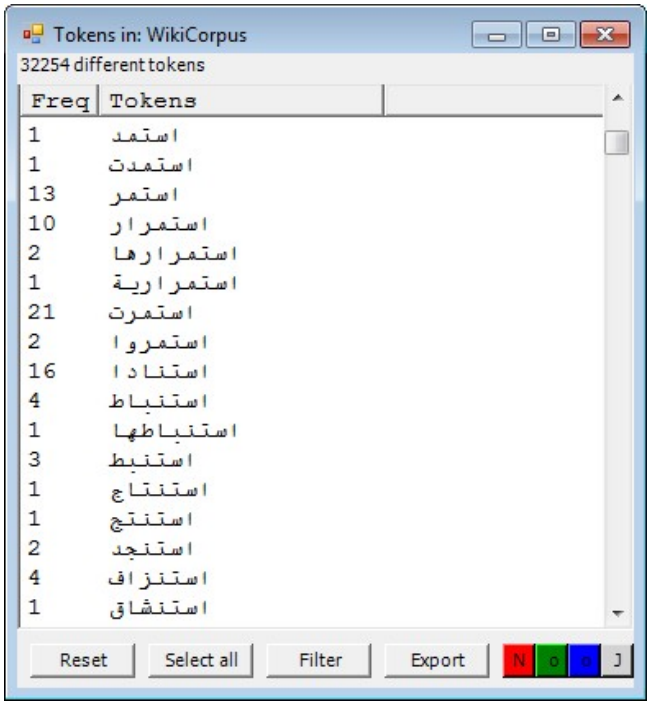
الشكل رقم (3)
الخُطُوات المنهجيَّة لبناء شبكة الكلمات العربيَّة المنشُودة

3.1 المرحلة الأولى: الإغناء

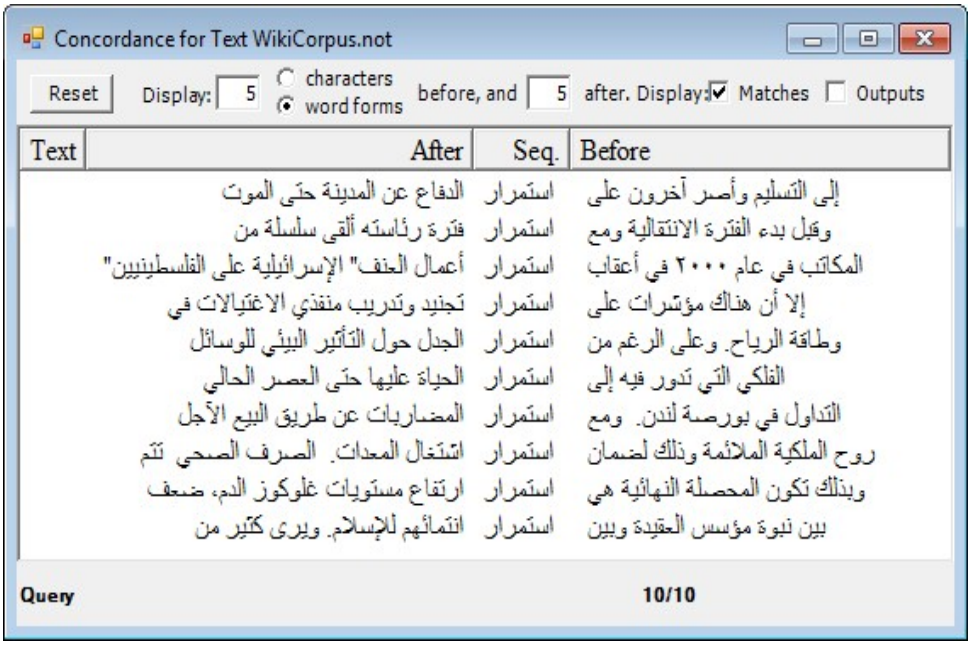
المادَّةُ الأساسيَّةُ الَّتي ينبغي أن تتضمَّنَها شبكةُ الكلمات المنشودة هي المباني (المُفرَدات) والمعاني (المفاهيم) والشَّواهد المُعبِّرة عن الاستخدام اللُّغَويّ للمباني والمعاني. ومادُمنا نلتمسُ شبكةً تعكسُ واقعَ اللُّغة العربيَّة المُستخدمة فعليًّا، فعلينا أن نستقي المادَّةَ من موردين رئيسَين، هما: المُدَوَّنة اللُّغَويَّة، وقاعدة البيانات المُعجميَّة.

3.1.1 المُدَوَّنة اللُّغَويَّة Linguistic Corpus

هي المصدر الأساسيّ للمُفرَدات في شبكة الكلمات؛ وهي أيضًا مصدرُ الشَّواهد الَّتي تُعَبِّرُ عن المُفرَدات ومعانيها [2]. وتقترحُ الدِّراسةُ أن تستمدَّ شبكةُ الكلمات مداخلها ووحداتها المُعجميَّة من مُدَوَّنةٍ لُغويَّةٍ للعربيَّة المُعاصرة، على أن نتوسَّعَ في مادَّتها مُستقبلًا حسبَ ما تقتضيه الحاجة. وحالَ إعداد المُدَوَّنة اللُّغَويَّة على الوجه المنشود، فعلينا أن نقومَ بمُعالجتها باستخدام آليَّة فهرسة النُّصُوص Text Indexer بهدف استخلاص المُفرَدات، على النَّحو المُبَيَّن في (الشَّكل 4)، ثُمَّ مُعالجتها باستخدام الكشَّاف السِّياقيّ Concordancer لاستخلاص الشَّواهد، على النَّحو المُبَيَّن في (الشَّكل 5).



الشكل رقم (4)
نموذج المُدَوَّنة اللُّغَويَّة المُفهرَسة آليًّا – من الموسُوعة الحُرَّة (ويكيبيديا)

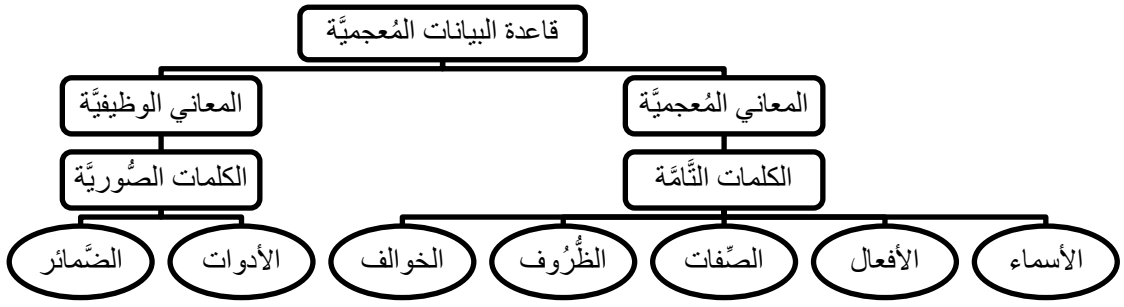


الشكل رقم (5)
نموذج المُدَوَّنة اللُّغَويَّة بعدَ مُعالجتها باستخدام الكشَّاف السِّياقيّ – من الموسُوعة الحُرَّة (ويكيبيديا)

3.1.2 قاعدة البيانات المُعجميَّة Lexical Database

تُصَنَّفُ كلماتُ اللُّغة – عُمومًا – إلى كلماتٍ تامَّة Full words وكلماتٍ صُوريَّة Form words. فالكلمات التَّامَّة هي الَّتي تحمل معنى تامًّا في ذاتها؛ ومصدرُها المُعجم Dictionary، وتُمَثِّلُها في اللُّغة العربيَّة – وفقَ التَّقسيم السُّباعيّ: الأسماء (كتاب، قَلَم، مدرَسة، ...) والأفعال (كَتَبَ، يفهم، يُغادِر، ...) والصِّفات (أخضَر، هادئ، سقيم،...) والظُّرُوف (تحت، أسفل، خلف،...) والخوالف (نِعمَ، بِئسَ، حبَّذا، ...)، أمَّا الكلمات الصُّوريَّة فهي الَّتي لا يُعرَف معناها إلاَّ بالنَّظر إلى الكلمات المُجاورة لها؛ ومصدرُها علم التَّركيب Syntax، وتُمَثِّلُها: الأدوات (مِنْ، ما، هل، ...)، والضَّمائر (نحن، أنتَ، هي، هُم، ...) [1، 5].

وإذا كانت المُدَوَّنةُ مصدرَ المباني في شبكة الكلمات، فإنَّ قاعدةَ البيانات المُعجميَّةَ هي مصدرُ المعاني. وتحقيقًا للغاية من شبكة الكلمات العربيَّة، فالمُفتَرَضُ أن تشتملَ قاعدةُ البيانات على المعاني المُعجميَّة Lexical Meanings الَّتي تُصاحبُ الكلمات التَّامَّة (الأسماءَ والأفعال والصِّفات والظُّروف والخوالف)، والمعاني الوظيفيَّة (النَّحويَّة)Grammatical Functions الَّتي تُصاحبُ الكلماتِ الصُّوريَّة (الأدوات والضَّمائر).



الشكل رقم (6)
مُحتوى قاعدة البيانات المُعجميَّة لشبكة الكلمات العربيَّة المنشُودة

وإثراءً لمُحتوى قاعدة البيانات المُعجميَّة، تقترحُ الدِّراسةُ أن تستمدَّ القاعدةُ مادَّتَها من معاجمَ عربيَّة، تعكسُ في مجموعها واقعَ اللُّغة العربيَّة المُعاصِرة؛ على أن تُزوَّدَ قاعدةُ البيانات – مُستقبلًا – بمادَّةٍ مُعجميَّةٍ مُمثِّلةٍ للعربيَّة في مَرحَلَتَيها القديمة والوسيطة، في ضوء ما تقتضيه الحاجةُ من شبكة الكلمات العربيَّة، مع الأخذ في الاعتبار أن يكونَ المعنى المُعجميُّ أو الوظيفيُّ المُصاحبُ للكلمات مُستخدمًا فعليًّا، على النَّحو الَّذي تُبينُ عنهُ المُدَوَّنةُ اللُّغَويَّةُ لشبكة الكلمات بما تحويه من المباني (المُفرَدات) باعتبارها تمثيلًا لواقع اللُّغة العربيَّة. ومن المعاجم الَّتي يُمكنُ الاعتمادُ عليها في بناء قاعدة البيانات المُعجميَّة:

- "مُعجَم اللُّغة العَرَبِيَّة الحديثة المكتُوبة Arabisches Wörterbuch für die Schriftsprache der Gegenwart" للمُستَشرِق الألمانِيّ هانز فِير Hans Wehr (1909-1981). وقد نَشَرَهُ في لَيبتسك بألمانيا عامَ 1952 [18].
- "المُعجَم الوسيط"، الَّذي شاركَ في تحريره نُخبةٌ من اللُّغَويِّينَ والمُعجَمِيِّينَ بإشراف مجمع اللُّغة العربِيَّة في القاهرة. وصدَرَ في طبعتِهِ الأولى عامَ 1960 [8].
- "مُعجم اللُّغة العربيَّة المُعاصرة"، الَّذي حرَّرَه المُعجميُّ المصريّ أحمد مختار عُمَر (1933-2003) بمُساعدة فريق عمل. وقد طُبِعَ في القاهرة عامَ 2008 [7].

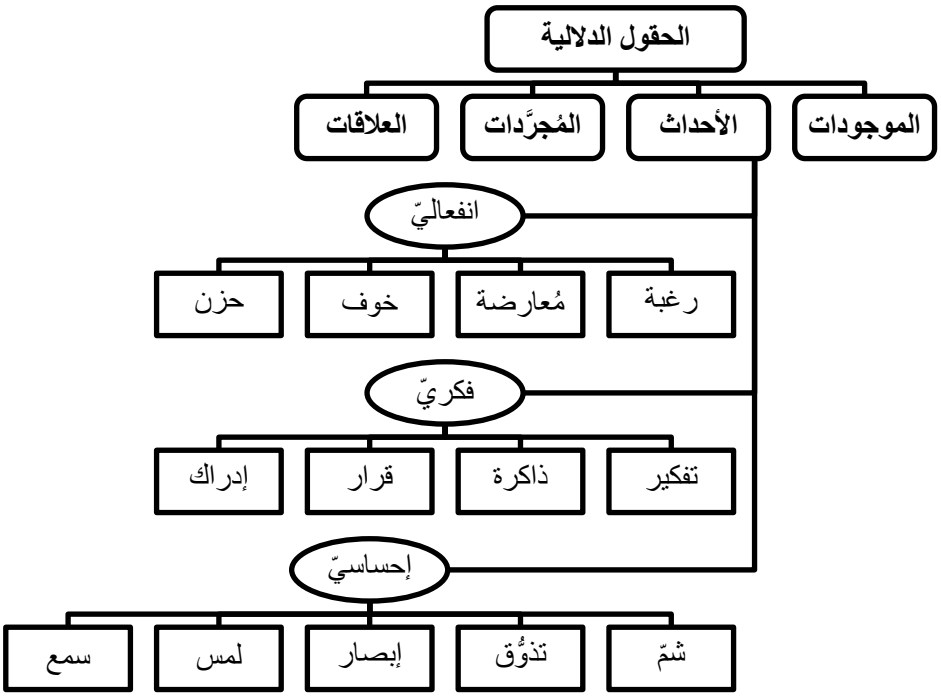
3.2 المرحلة الثَّانية: الرَّبط بينَ البيانات

تُعنى هذه المرحلةُ ببناء الشَّبكة العلائقِيَّة Relational Network بين المباني والمعاني في شبكة الكلمات العربيَّة المنشودة. والغايةُ من الشَّبكة العلائقيَّة إتاحةُ شبكة الكلمات في صُورةٍ يُستفادُ منها في جوانب المُعالجة الدِّلاليَّة، حيثُ تُصنَّفُ المُفرَداتُ ومعانيها في حُقُولٍ دلاليَّةٍ هرميَّةٍ، وترتبطُ فيما بينها بعلاقاتٍ دلاليَّة. وبيانُ ذلكَ على النَّحو الآتي.

3.2.1 الحُقُول الدِّلاليَّة Semantic Fields

يُعبِّرُ الحقلُ الدِّلاليُّ عن مجموعةٍ من الكلمات(المُفرَدات) الَّتي يَضُمُّها مجالٌ مَعرِفيٌّ مُعَيَّن، وترتبطُ هذه الكلماتُ فيما بينها بعلاقة دلاليَّة، تُحَدِّدُها طبيعةُ الحقل الَّذي تنتمي إليه. وقد ظهرَت العديدُ من المُحاولات لتصنيف كلمات اللُّغات الطَّبيعيَّة في حُقُولٍ دلاليَّة؛ لعلَّ أبرزَها مُحاولة اللُّغَويّ المجريّ "ستيفن أولمان Stephen Ullmann" (1914-1976) في كتابه (المعنى والأسلوب Meaning and style) [17]، ومُحاولة الألمانيَّين رودولف هالينج Rudolf Hallig (1902-1964) وفالتر فون فارتبورج Walther von Wartburg (1888-1971) في كتاب (الإطار المفاهيميّ المُعجميّ: مُحاولة تجريبيَّة Begriffssystem als Grundlage für die Lexikographie: Versuch eines Ordnungsschemas) [13]، ومُحاولة الأمريكيّ يوجين نايدا Eugene Nida (1914-2011) والجنوب إفريقيّ يوهانس لو Johannes P. Louw (1932-2011) في (المُعجم اليُونانيّ-الإنجليزيّ للعهد الجديد على أساس المجالات الدِّلاليَّة Greek-English lexicon of the New Testament based on semantic domains) [14].

وتقترحُ الدِّراسةُ تصنيفَ مُحتوى شبكة الكلمات المنشُودة وفقَ المنهج الَّذي سارَ عليه [نايدا & لو] في المُعجم اليُونانيّ-الإنجليزيّ، حيثُ أمكنَ تطبيقُهُ فعليًّا على مُفرَدات العهد الجديد الَّتي صُنِّفَت إلى أربعة حُقُولٍ دلاليَّة رئيسةٍ، هي: الموجودات Entities، والأحداث Events، والمُجرَّدات Abstracts، والعلاقات Relations. وتتفَرَّعُ هذه الحُقُول الرَّئيسة لتشمل حُقُولاً فرعيَّة، تتفرَّعُ بدورها إلى حُقُولٍ فرعيَّةٍ أخرى ... وهكذا. ويُوَضِّحُ (الشَّكل 7) نموذجًا لبعض الحُقُول الدِّلاليَّة المُتفرِّعة عن حقل (الأحداث) في هذا المُعجم.



الشكل رقم (7)
نموذج من الحقل الدِّلاليّ (الأحداث) – عن (المُعجم اليُونانيّ-الإنجليزيّ للعهد الجديد)

3.2.2 العلاقات الدِّلاليَّة Semantic Relations

إذا كانت الحُقُولُ الدِّلاليَّة وعاءً للمُفرَدات المُتجانسة، فإنَّ العلاقات الدِّلاليَّة هي وسيلةُ الرَّبط بينَ هذه المُفرَدات. والمُفتَرَضُ أن تربطَ العلاقاتُ الدِّلاليَّة بينَ جميع المُفردات (الوحدات اللُّغويَّة) في شبكة الكلمات، بحيثُ لا تبقى مُفرَدةٌ واحدةٌ إلَّا وقد ارتبَطَت بغيرها بعلاقةٍ دلاليَّةٍ أو أكثر. وتحقيقًا لهذه الغاية، تقترحُ الدِّراسةُ تصنيفَ العلاقات الدِّلاليَّة في شبكة الكلمات إلى: العلاقات الأفقيَّة والعلاقات الرَّأسيَّة. وتمثيلُ ذلك على النَّحو الآتي.

3.2.2.1 العلاقات الأفقيَّة Horizontal Relations

هي العلاقاتُ الدِّلاليَّةُ الَّتي تربطُ بينَ المُفرَدات في مُستوى دلاليٍّ واحدٍ، ومثالُها العلاقات في (الجدول 1).

الجدول رقم (1)
مُقتَرَح العلاقات الدِّلاليَّة الأفقيَّة في شبكة الكلمات العربيَّة

م	العلاقة	المفهوم	المثال
1	التَّرادُف *Synonymy*	إذا كانَ (أ) يشتمل على (ب)، و (ب) يشتمل على (أ)، فإنَّ (أ) و (ب) مُترادفان.	العام = السَّنة = الحول
2	التَّضادّ *Antonymy*	إذا كانَ (أ) ضدّ (ب)، و (ب) ضدّ (أ)، فإنَّ (أ) و (ب) مُتضادَّان.	الصِّدق × الكذب
3	التَّنافُر *Incompatibility*	إذا كانَ (أ) يشتمل على (ب)، و (ج)، وكانَ (ب) ≠ (ج)، فإنَّ (ب) و (ج) مُتنافران.	رَجُل ≠ طِفل (إنسان)
4	الاشتراك اللَّفظيّ [الهومونيمي] *Homonymy*	إذا كانَ (أ) = (ب) لفظًا، وكانَ (أ) ضدّ (ب) معنًى، فإنَّ (أ) أو (ب) مُشتَرَك لفظيّ هومونيمي.	مُختار [الفاعل × المفعول]
5	الاشتراك اللَّفظيّ [البوليزيمي] *Polysemy*	إذا كانَ (أ) = (ب) أو (ج)، وكانَ (ب) ضدّ (ج)، فإنَّ (أ) مُشتَرَك لفظيّ بوليزيميّ.	الصَّريم = اللَّيل (أو) النَّهار اللَّيل × النَّهار

3.2.2.2 العلاقات الرَّأسيَّة Vertical Relations

هي العلاقاتُ الدِّلاليَّةُ الَّتي تربطُ بينَ المُفرَدات في مُستوياتٍ دلاليَّةٍ مُتتالية، ومثالُها العلاقات في (الجدول 2).

الجدول رقم (2)
مُقترَح العلاقات الدّلاليَّة الرَّأسيَّة في شبكة الكلمات العربيَّة

م	العلاقة	المفهوم	المثال
1	النّوعِيَّة *Hyponyms*	إذا كانَ (أ) نوعًا من (ب)، فالعلاقة بينَ (أ) و (ب) نوعيَّة.	الأحمر – اللَّون
2	الكُلِّيَّة *Holonym*	إذا كانَ (أ) كُلًّا من (ب)، فالعلاقة بينَ (أ) و (ب) كُلِّيَّة.	الحِصان – الذَّيل
3	الجُزئِيَّة *Meronym*	إذا كانَ (أ) جُزءًا من (ب)، فالعلاقة بينَ (أ) و (ب) جُزئيَّة.	الذَّيل – الحصان
4	التَّلازُمِيَّة *Troponym*	إذا كانَ (أ) يُؤدِّي إلى (ب)، وكان (ب) لا يتمُّ إلّا عن طريق (أ)، فإنَّ (أ) و (ب) مُتلازمان.	يُتَهتِه / يتكلَّم
5	الاستلزاميَّة *Entailment*	إذا كانَ (أ) لا يتمُّ إلّا عن طريق (ب)، فإنَّ (أ) يستلزم (ب).	ينام / يغفو

ولتيسير الحُصُول على العلاقات الدّلاليَّة بينَ المُفرَدات، تقترحُ الدِّراسةُ إخضاع المُدَوَّنة اللُّغَويَّة الَّتي تستمدُّ منها شبكةُ الكلمات وحداتها لتقنيات (النَّحو العدديّ N-Gram) بهدف استخلاص النَّماذج اللُّغَويَّة Language models الَّتي تتحدَّدُ من خلالها العلاقاتُ الدّلاليَّة. وقد قامَ الباحثُ بمحاولةٍ لاستخلاص النَّماذج اللُّغَويَّة من مُدَوَّنةٍ لُغويَّةٍ مُصَغَّرةٍ، تحتوي على 1000 مقالةٍ عربيَّة مُستمدَّةٍ من الموسُوعة الحُرَّة (ويكيبيديا Wikipedia) الَّتي تعكسُ صُورةً للُّغة العربيَّة المُعاصرة. وأمكنَ من خلال المُعالجة استخلاصُ عدد من النَّماذج اللُّغَويَّة الَّتي يُمكنُ توظيفُها في تعيين علاقة (التَّرادُف) باعتبارها إحدى العلاقات الدّلاليَّة الأفقيَّة على النَّحو الوارد في (الجدول 3).

الجدول رقم (3)
النَّماذج اللُّغَويَّة لعلاقة التَّرادُف – من الموسُوعة الحُرَّة (ويكيبيديا)

م	السَّابق	الرَّابط	اللّاحق
1	أما الفقه الإسلامي فهو	عبارة عن	العلم بالأحكام الشرعية العملية المستنبطة
2	وهي سيدة الوردية المقدسة، والوردية	عبارة عن	صلوات تأملية في مسيرة المسيح
3	على أن هذا الكوكب هو	عبارة عن	بيئة متكاملة تتطلب إشراف الإنسان
4	هذه الألواح التكتونية	عبارة عن	أجزاء صلبة تتحرك مع بعضها
5	... في النظام الشمسي من سديم شمسي	عبارة عن	كتلة قرصية الشكل من الغبار
6	... الذي تتدفق فيه الأيونات، وهي	عبارة عن	ذرات مشحونة
7	... كما اكتشف الكهرومغناطيسية، وهي	عبارة عن	تفاعل يحدث بين الكهرباء والمغناطيسات
8	التناظر الهيدروليكي	عبارة عن	التناظر بين تدفق الماء والتيار الكهربائي
9	الشحنة الكهربائية	عبارة عن	خاصية لبعض الجسيمات دون الذرية
10	الدائرة الكهربائية	عبارة عن	ترابط بين المكونات الكهربائية

كما أمكنَ استخلاصُ عدد من النَّماذج اللُّغَويَّة الَّتي يُمكنُ توظيفُها في تعيين علاقة (الجُزئيَّة) باعتبارها إحدى العلاقات الدّلاليَّة الرَّأسيَّة على النَّحو الوارد في (الجدول 4).

الجدول رقم (4)
النَّماذج اللُّغويَّة لعلاقة الجزئيَّة – من الموسُوعة الحُرَّة (ويكيبيديا)

م	السَّابق	الرَّابط	اللّاحق
1	تمثل الطيور أهم	جزء من	الثروة الحيوانية المغربية
2	... الضفة والقدس والقطاع على أنهم	جزء من	الشعب الفلسطيني وليسوا أردنيين
3	الزكاة: وهي	جزء من	المال يدفعه الأغنياء من المسلمين
4	إلى ذلك أن العهد القديم	جزء من	وحي الله
5	الاتحاديون يريدون أن تبقى أيرلندا الشمالية	جزء من	المملكة المتحدة
6	تقع الخرطوم في حوض الخرطوم الذي يشكل	جزء من	حوض النيل الأزرق
7	أصبحت أراضي كرواتيا الحالية	جزء من	الإمبراطورية الرومانية
8	كنيسة التوحيدية الأثيوبية الأرثذوكسية هي	جزء من	كنائس الأرثوذكس المشرقين
9	ويعتبر القدس الشرقية	جزء من	الأراضي الفلسطينية
10	الأزور وماديرا في المحيط الأطلسي هي أيضا	جزء من	البرتغال

3.3 المرحلة الثَّالثة: هيكلة شبكة الكلمات

تُعنى هذه المرحلةُ ببناء هيكل شبكة الكلمات على النَّحو الَّذي يُحَقِّقُ الغايةَ المنشودةَ منها. ومادُمنا نلتمسُ توظيفَ الشَّبكة في أغراض صناعة المُعجم العربيّ، فالدِّراسةُ تقترحُ هيكلةَ شبكة الكلمات المنشودة في مُستوَيَين رَئيسَين، هما: (مُستوى الهيكلة المُعجميَّة) الَّذي يسمحُ بتوظيف الشَّبكة في صناعة مُعجمات الألفاظ والمباني، و (مُستوى الهيكلة الدِّلاليَّة) الَّذي يسمحُ بتوظيف الشَّبكة في صناعة مُعجمات الموضُوعات والمعاني. وبيانُ ذلك على النَّحو الآتي:

3.3.1 الهيكلة المُعجميَّة

تقومُ الهيكلةُ المُعجميَّةُ لشبكة الكلمات العربيَّة المنشُودة على مجموعةٍ من المُعطيات، تُجملُها الدِّراسةُ في النِّقاط الآتية:

- توزيع أقسام الكلام السَّبعة (الاسم، والفعل، والصِّفة، والظَّرف، والخالفة، والأداة، والضَّمير) على ثلاثة أقسامٍ رئيسةٍ، هي: (الاسم، والفعل، والأداة) وفقَ ما يتناسبُ مع معايير الصِّناعة المُعجميَّة العربيَّة.

- ترتيب المداخل والوحدات المُعجميَّة ألفبائيًّا – على نسَق المُعجمات العربيَّة المُعاصِرةِ – مع مُراعاة تقديم الأفعال على الأسماء والأدوات، وتقديم الثُّلاثيّ منها على الرُّباعِيّ والخُماسيّ والسُّداسيّ، وتقديم المُجَرَّد منها على المَزِيد، وتقديم اللاَّزِم على المُتَعَدِّيّ.

- نمذجة المعاني المُعجميَّة للوحَدات بشكلٍ مُنتظمٍ، لتيسير استدعاء المُفرَدة الواحدة الَّتي يُصاحبُها أكثرُ من معنى أو المُفرَدات المُتعدِّدة الَّتي تشتركُ في معنى واحدٍ – حالَ الحاجة إلى ذلك.

- الاقتصار على المعاني الوظيفيَّة السَّماعيَّة على المُستوَيَين البنَويّ والتَّركيبيّ (كتعيين أبواب الفعل الثُّلاثيّ وجُمُوع الأسماء وصِيَغ التَّذكير والتَّأنيث، وتمييز اللَّازم والمُتعدِّي من الأفعال، وتمييز علامات أقسام الكلام ونحو ذلك)، دونَ التَّعَرُّض للجانب القِياسِيّ الَّذي تُعنَى بِه كُتُب النَّحو العَرَبِيّ.

- مُراعاة وُضُوح الشَّواهد المُعجميَّة وتمثيلها للمباني والمعاني على نحوٍ دقيقٍ مُستمدٍّ من واقع اللُّغة المُستخدمَة فعليًّا، مع ربط الشَّواهد بمُستويات الاستعمال المُعبِّرة عن درجة شُيُوع المُفرَدات وإطارها الاجتماعيّ واعتبارات الزَّمان والمكان، ثُمَّ توثيق ذلك من المُدَوَّنة اللُّغَويَّة للشَّبكة باعتبارِها المصدرَ الأساسيَّ للشَّواهد.

3.3.2 الهيكلة الدِّلاليَّة

وتقومُ الهيكلةُ الدِّلاليَّة للشَّبكة المنشُودة – كذلك – على مجموعةٍ من المُعطيات، تُجملُها الدِّراسةُ في النِّقاط الآتية:

- التَّمييز بينَ الحُقُول الدِّلاليَّة الَّتي تحوي الكلمات التَّامَّة، وفقَ المنهج الَّذي تسيرُ عليه شبكةُ الكلمات في تعيين حُقُولها، مع إضافة حقلٍ خاصٍّ يحوي (الكلمات الصُّوريَّة) باعتبارها مُفرَدات لا تحملُ معنى في ذاتِها.

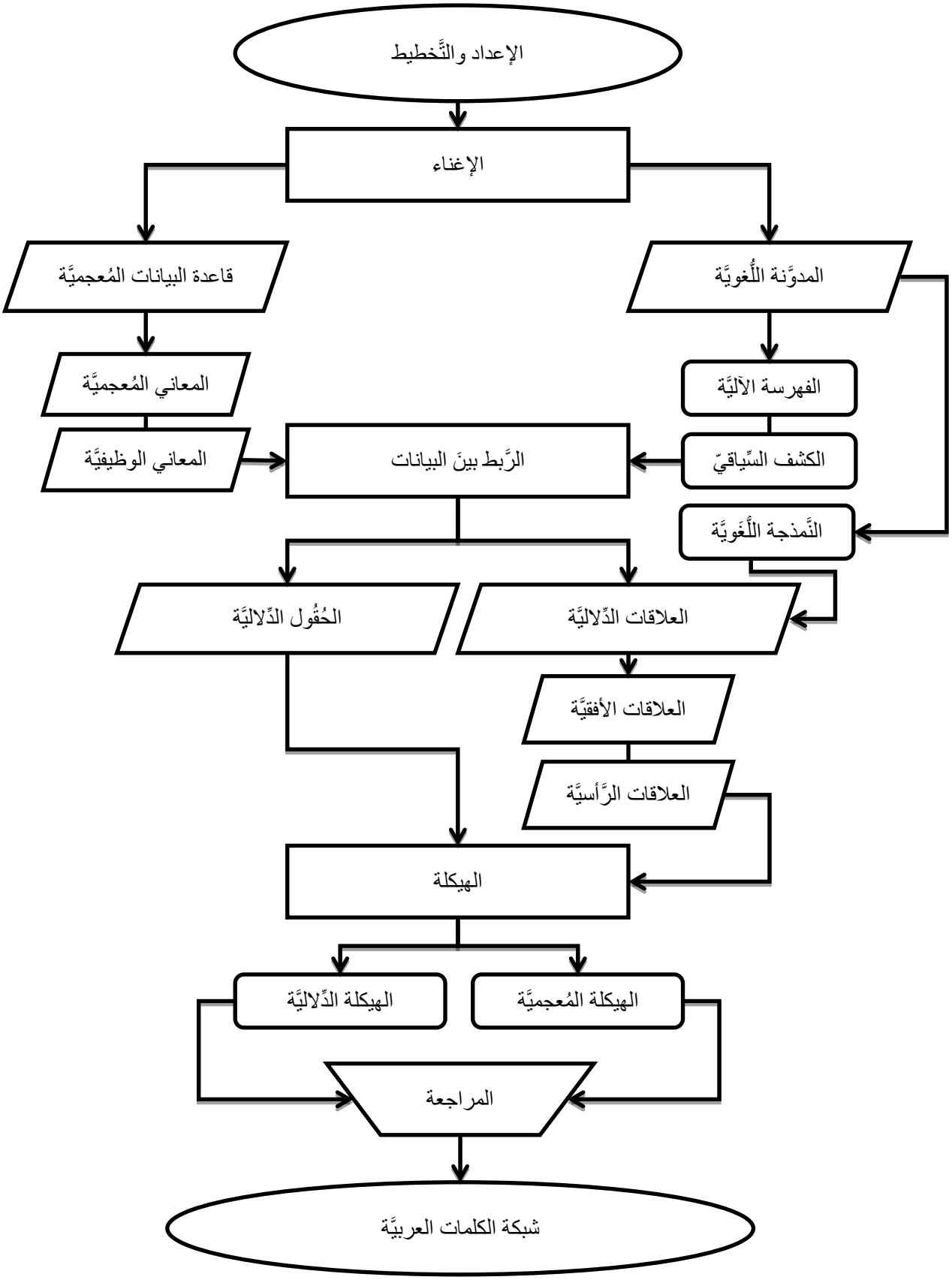
- توزيع المُفرَدات في بناءٍ شجريّ (أو هرميّ) للحُقُول الدِّلاليَّة الَّتي ستحويها المُفرَداتُ، مع تحديد مُستوى رئيسٍ يُمثِّلُ الحُقُول الدِّلاليَّةَ الرئيسة، ومُستوياتٍ ثانويَّة تُمثِّلُ الحُقُولَ الدِّلاليَّةَ الفرعيَّة.

- إدراج كُلِّ مُفرَدةٍ على حِدةٍ في حقلٍ دلاليٍّ مُعَيَّنٍ، دونَ إدراجها في حُقُولٍ أخرى، بحيثُ يبدوُ كلُّ حقلٍ دلاليٍّ مُستقلًّا وجامعًا لمُفرَداتٍ لا تتكرَّرُ في الحُقُول الأخرى، على النَّحو الَّذي تسيرُ عليه المداخلُ المُعجميَّة.

- الجمع بينَ المُفرَدة الواحدة ومعانيها المُتعدِّدة في الحقل الدِّلاليّ الَّذي تنتمي إليه، وربط المُفرَدات الَّتي تشتركُ في معنى واحدٍ عن طريق العلاقات الدِّلاليَّة المُدرَجة في شبكة الكلمات.

- مُراعاة التَّوازي بينَ مُستويات الحُقُول الدِّلاليَّة الَّتي ترتبطُ المُفرَداتُ فيها بعلاقاتٍ دلاليَّةٍ أفقيَّة (مثل: التَّرادُف)، ومُراعاة التَّوالي بينَ مُستويات الحُقُول الدِّلاليَّة الَّتي ترتبطُ المُفرَداتُ فيها بعلاقاتٍ رأسيَّة (مثل: النَّوعيَّة).

- التَّمييز بينَ العلاقات الدِّلاليَّة الأفقيَّة والرَّأسيَّة، ومُراعاة تعدُّد المُفرَدات الَّتي تشتركُ في علاقةٍ دلاليَّةٍ واحدة، مع إحداث التَّجانُس بينَ أقسام الكلام بحيثُ تشتركُ المُفرَداتُ الَّتي ترتبطُ بعلاقةٍ مُعَيَّنةٍ في نفس قسم الكلام.



الشكل رقم (8)
مُخطَّط انسيابيّ لمنهجيَّة بناء شبكة للكلمات العربيَّة

4 توظيف شبكة الكلمات العربيَّة في أغراض الصِّناعة المُعجَميَّة

تُعاني الصِّناعةُ المُعجميَّةُ العربيَّةُ قُصُورًا في العديد من جوانبها. ومن أسباب هذا القُصُور اعتمادُ الصِّناعة على رؤية المُعجَميِّينَ الَّتي تبدو خاصَّةً بهم في كثير من الأحيان، بصرف النَّظر عن المناهج والمعايير الَّتي تضبطُ الصِّناعةَ المُعجميَّة. وتكونُ الحصيلةُ حينئذٍ أن يتغيَّرَ منهجُ المُعجم بتغيُّر مُحرِّره أو تعاقُب الأجيال عليه. ومن ناحيةٍ أخرى، تستقي الصِّناعةُ المُعجميَّةُ العربيَّةُ المُعاصرةُ مادَّتَها – غالبًا – من معاجمَ أخرى بعدَ تهذيب مُحتواها بما يتناسبُ مع الغاية من عملهم. وفي أحيانٍ قليلة، تُضافُ إليها بعضُ كُتُب التُّراث ودواوين الشِّعر العربيّ بهدف إثراء المادَّة المُعجميَّة وتوفير الشَّواهد الَّتي تُعَبِّرُ عن الاستعمال اللُّغَويّ للمُفرَدات ومعانيها. وعليه، فإنَّنا نُلاحظُ إهمالَ الصِّناعةِ المُعجميَّةَ العربيَّةَ للموارد الَّتي يُمكنُ الإفادةُ منها في توفير الوقت والجهد وإخراج العمل المُعجميّ على وجهه الَّذي تتحقَّقُ معهُ الغايةُ منه.

والواقعُ أنَّ الصِّناعةَ المُعجميَّةَ في عُمُومها تفرضُ أن تتَّفقَ المُعجماتُ – على اختلاف أنواعِها – في بعض المعلومات المُعجميَّة وتختلفَ في بعضِها الآخر، وفقَ ما تُحدِّدُهُ طبيعةُ المُعجم وما تقتضيه الغايةُ منه. وعلى سبيل المثال، يتَّفقُ الإطارُ العامُّ لمُعجمٍ موضُوعيٍّ للأطفال مع مُعجمٍ موضُوعيٍّ يُقَدَّمُ للشَّباب، وإن اختلفَ المُعجمان في حجم المادَّة المُتَضَمَّنة أو موسُوعيَّة المعلومات الَّتي تحويها حُقُولُ المُعجم. والحالُ كذلك في مُعجمات الألفاظ، حيثُ يتَّفقُ مُحتواها في معلومات الهجاء والنُّطق والمداخل والوحدات والمعاني الوظيفيَّة وبعض المعاني المُعجميَّة وغير ذلك من المعلومات المُعجميَّة. نُلاحظُ مثلًا اتِّفاقَ مادَّةَ المُعجمَين (الوجيز والوسيط) الَّذين أنجزَهما مجمعُ اللُّغة العربيَّة بالقاهرة، وإن كانت مُوجزةً في (الوجيز) المصنوع لإفادة الطُّلَّاب في مراحل التَّعليم الأساسيّ والثَّانويّ، وأكثرَ تفصيلًا في (الوسيط) الَّذي أريدَ به مواءمة رُوح العصر الحديث.

في ضوء هذه المُعطيات، يُمكنُ القَولُ إنَّ الحاجةَ تدعو إلى بناءَ شبكةٍ للكلمات العربيَّة، لتكونَ موردًا لُغويًّا وافيًا يُستفادُ منهُ في أغراض الصِّناعة المُعجميَّة، بما يتضمَّنُهُ هذا المورِدُ من المباني والمعاني والشَّواهد والمعلومات المُعجميَّة الأخرى. ذلك أنَّ بناءَ شبكة الكلمات المنشودة بما يتناسبُ مع طبيعة اللُّغة العربيَّة وأنظمتها اللُّغَويَّة سيُثمرُ بالضَّرورةِ عن قاعدة بياناتٍ مُعجميَّةٍ دلاليَّةٍ ضخمة تتشكَّلُ في هيكلٍ معلوماتيٍّ كبير، يُمكنُ توظيفُهُ في صناعة سلسلةٍ من معاجم الألفاظ المرحليَّة والتَّعليميَّة ومعاجم الموضوعات لمختلف الأعمار ومعاجم مُتخصِّصة في مُصطلحات العُلُوم والفُنون والآداب ومكانز لُغويَّة، وغير ذلك من أشكال الصِّناعة المُعجميَّة، كما يُمكنُ توظيفُهُ في بناء أنطولوجيا معرفيَّة وفي تطبيقات حَوسبة اللُّغة وبناء أدوات مُعالجتِها آليًّا.

ويُوضِّحُ (الجدول 5) تصوُّرًا لكيفيَّة توظيف شبكة الكلمات العربيَّة في أغراض الصِّناعة المُعجَميَّة.

الجدول رقم (5)
بيان توضيحيّ لكيفيَّة توظيف شبكة الكلمات في صناعة المُعجم العربيّ

مادَّة شبكة الكلمات	معاجم الألفاظ	معاجم الموضُوعات	معاجم المُصطلحات	المكانز اللُّغَويَّة
المداخل والوحدات	+	+	+	+
معلومات الهجاء والنَّطق	+	+		
المعاني الوظيفيَّة	+	+		+
المعاني المُعجميَّة	+	+	+	+
الشّواهد المُعجميَّة	+			
معلومات التّأثيل والتّأريخ	+			
مُستويات الاستعمال	+	+		+
معلومات التّكرار	+			+
الحُقُول الدِّلاليَّة		+	+	+
العلاقات الدِّلاليَّة		+		+

5 نتائج الدِّراسة

1. اقتَرَحَت الدِّراسةُ منهجِيَّةً لبناء شبكة الكلمات العربيَّة المنشُودة عبرَ ثلاث مراحلَ أساسيَّة، حيثُ تُعنى المرحلةُ الأولى بتوفير موارد شبكة الكلمات المنشودة، وتُعنى المرحلةُ الثَّانيةُ بالرَّبط بينَ الوحدات اللُّغَويَّة المُتَضَمَّنة في شبكة الكلمات باستخدام الحُقُول والعلاقات الدِّلاليَّة، وتُعنى المرحلةُ الثَّالثةُ والأخيرةُ ببناء هيكل شبكة الكلمات وفقَ ما يُحقِّقُ غايةَ الإفادة من الشَّبكة في أغراض الصِّناعة المُعجميَّة العربيَّة.
2. اقتَرَحَت الدِّراسةُ أن تستمدَّ شبكةُ الكلمات مبانيها وشواهدِها من مُدَوَّنةٍ لُغويَّةٍ للعربيَّة المُعاصرة، بعدَ إخضاعِها للمُعالجة الآليَّة، على أن نتوسَّعَ في مادَّتها مُستقبلًا حسبَ ما تقتضيه الحاجة؛ وأن تستمدَّ المعاني المُعجميَّة والوظيفيَّة من قاعدة بياناتٍ مُعجميَّةٍ تُصنَعُ لغاية إثراء مُحتوى شبكة الكلمات.
3. اقتَرَحَت الدِّراسةُ تصنيفَ مُحتوى شبكة الكلمات المنشُودة إلى أربعة حُقُولٍ دلاليَّةٍ رئيسةٍ، هي: الموجودات والأحداث والمُجرَّدات والعلاقات، تتفرَّعُ عنها حُقُولٌ أخرى، وفقَ المنهج الَّذي سارَ عليه (نايدا & لو) في (المُعجم اليُونانيّ-الإنجليزيّ للعهد الجديد على أساس المجالات الدِّلاليَّة).
4. اقتَرَحَت الدِّراسةُ تصنيفَ العلاقات الدِّلاليَّة في شبكة الكلمات إلى: العلاقات الأفقيَّة، ومثَّلَ لها الباحثُ بعلاقات (التَّرادُف، والتَّضادّ، والتَّنافُر، والاشتراك اللَّفظيّ الهُومونيمي، والاشتراك اللَّفظيّ البوليزيمي)، والعلاقات الرَّأسيَّة، ومثَّلَ لها الباحثُ بعلاقات (النَّوعيَّة، والكُلِّيَّة، والجُزئيَّة، والتَّلازُمِيَّة، والاستلزاميَّة). واقتَرَحَت الدِّراسةُ تعيينَ العلاقات الدِّلاليَّة باستخدام تقنيات (النَّحو العدديّ N-Gram) بهدف استخلاص النَّماذج اللُّغَويَّة. وقامَ الباحثُ بمحاولةٍ لاستخلاص النَّماذج اللُّغَويَّة من مُدَوَّنةٍ لُغويَّةٍ للعربيَّة المُعاصرة، فأمكنَ استخلاصُ عدد من النَّماذج اللُّغَويَّة الَّتي يُمكنُ توظيفُها في تعيين علاقاتٍ دلاليَّةٍ أفقيَّةٍ ورأسيَّة.

5. اقتَرَحَت الدِّراسةُ هيكلةَ شبكة الكلمات المنشودة في مُستوَيَين رَئيسَين، هما: (مُستوى الهيكلة المُعجميَّة) الَّذي يسمحُ بتوظيف الشَّبكة في صناعة مُعجمات الألفاظ والمباني، و (مُستوى الهيكلة الدِّلاليَّة) الَّذي يسمحُ بتوظيف الشَّبكة في صناعة مُعجمات الموضُوعات والمعاني.
6. قدَّمَ الباحثُ تصوُّرًا لكيفيَّة توظيف شبكة الكلمات العربيَّة في أغراض الصِّناعة المُعجَميَّة، وأبانَ عن إمكانيَّة الإفادة من مُحتوى الشَّبكة في صناعة معاجم الألفاظ المرحليَّة والتَّعليميَّة ومعاجم الموضوعات والمعاجم المُتخصِّصة في مُصطلحات العُلُوم والفُنون والآداب والمكانز اللُّغويَّة.

6 الخُلاصة

عَرَفَت اللُّغةُ العربيَّةُ طريقَها إلى شبكة الكلمات في مطلع الألفيَّة الثَّالثة، وظهرَت عدَّةُ مشروعاتٍ لبناء وتطوير شبكاتٍ للكلمات العربيَّة، إلَّا أنَّ أيًّا منها لم يكتمل على الصُّورة الَّتي تسمحُ بالإفادة منها في الصِّناعة المُعجميَّة. والواقعُ أنَّ ثمَّةَ إشكالاتٍ تواجهُ بناءَ شبكة للكلمات العربيَّة؛ منها: طبيعةُ النِّظام المُعجميّ العربيّ الَّذي يقتضي التَّمييزَ بينَ المداخل والوحدات المُعجميَّة ومُراعاة تقسيم الكلام، وطبيعة اللُّغة العربيَّة في مبانيها الَّتي تتعدَّدُ أشكالُها باختلاف حركات ضبطها، وطبيعة اللُّغة العربيَّة في معانيها الَّتي تقتضي مُراعاةَ أثر البيئة في تعدُّد الدِّلالات. وانطلاقًا من طبيعة اللُّغة العربيَّة ونظامها المُعجميّ من ناحية، واستقراء التَّجارب والمُحاولات السَّابقة لبناء شبكاتٍ لكلمات اللُّغات الإنسانيَّة من ناحيةٍ أخرى، تقترحُ الدِّراسةُ منهجيَّةً لبناء وتطوير شبكةٍ للكلمات العربيَّة لأغراض الصِّناعة المُعجميَّة عبرَ ثلاث مراحلَ أساسيَّة، هي: الإغناء، والرَّبط بينَ البيانات، وهيكلة شبكة الكلمات. وقد فصَّلَ الباحثُ القولَ في المراحل الثَّلاثة في صُورة خُطواتٍ منهجيَّةٍ تكونُ مُحصِّلتُها شبكة الكلمات المنشودة، كما قدَّمَ تصوُّرًا لكيفيَّة توظيف هذه الشَّبكة في أغراض الصِّناعة المُعجَميَّة العربيَّة.

قائمة المراجع

[1] حَسَّان (تَمَّام): اللُّغة العَرَبِيَّة "مَعناها وَمَبناها"، الهَيئة المِصرِيَّة العامَّة للكِتاب، القاهِرة، ط2، 1979.

[2] السَّعيد (المُعتزّ بالله): المدوَّنات اللُّغَوِيَّة، ضِمنَ كتاب (مُقَدِّمة في حوسَبة اللُّغة العربيَّة)، مجموعة من المؤلِّفين، تحرير: مُحسن رَشوان، والمعتزّ بالله السَّعيد، قَيد النَّشر بمدينة الملك عبد العزيز للعُلُوم والتِّقنية، الرِّياض، الجزء الأوَّل، 2015م.

[3] السَّعيد (المُعتزّبالله): شبكة الكلمات العَرَبِيَّة "المنهجِيَّة والتَّطبيق"، مجلَّة مجمَع اللُّغة العَرَبيَّة بدمشق، المُجلَّد 87، الجُزء الثَّاني، 2012.

[4] السَّعيد (المُعتزّبالله): شبكة الكلمات العَرَبِيَّة وتطبيقاتها في صناعة المُعجم العربيّ المُعاصِر، أعمال النَّدوة الدَّولِيَّة للُّغة والهُوِيَّة العربيَّة، كُلِّيَّة الآداب والعُلُوم في جامعة قَطَر، الدَّوحة، أبريل 2012.

[5] عبد الدَّايم (مُحمَّد): العربيَّة من منظور المنهج التَّوزيعيّ، مركز النَّشر العلميّ، جامعة الملك عبد العزيز، جدَّة، ط1، 2012.

[6] عطيَّة (مُحمَّد): دراسة حولَ واقع جهود البحث والتَّطوير في مَيدان التَّحليل الدِّلاليّ للُّغة العربيَّة، اجتماع خُبراء الشَّبكات الدِّلاليَّة للمُفردات العربيَّة، المنظَّمة العربيَّة للتَّربية والثَّقافة والعُلُوم، تونس، 2011.

[7] عُمَر (أحمد مُختار): مُعجم اللُّغة العربيَّة المُعاصرة، عالَم الكُتُب، القاهرة، 2008.

[8] مُصطفى (إبراهيم)، وآخرين: المُعجَم الوسيط، مَجمع اللُّغة العَرَبِيَّة، ط3، 1998.

[9] Abouenour, L. & Bouzoubaa, K. & Rosso,P. (2013). On the evaluation and improvement of arabic wordnet coverage and usability. Language Resources and Evaluation. September 2013, Volume 47 (3), pp 891-917.

[10] 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, GWC 2006, South Jeju Island, Korea, January 22-26, 2006.

[11] Buitelaar, P. & Cimiano, P. (2014). Towards the Multilingual Semantic Web: Principles, Methods and Applications. Springer.

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

[13] Hallig, R. & Wartburg, W. (1952). Begriffssystem als Grundlage für die Lexikographie. Begriffssystem. Akademie-Verlag.

[14] Louw, J. & Nida, E. (1989). Greek-English Lexicon of the New Testament based on Semantic Domains. Volume 1. United Bible Societies.

[15] Mubarak, H. (2011). Arabic WordNet Enrichment Using Sakhr Semantic Features and Lexical Database. Arabic WordNet meeting, (ALECSO) 26-28, July, 2011, Tunisia.

[16] Poli, R. & Seibt, J. (2010). Theory and Applications of Ontology: Philosophical Perspectives. Springer Science & Business Media.

[17] Ullmann, S. (1973). Meaning and style. Barnes & Noble Books.

[18] Wehr, H. (2011). Arabisches Wörterbuch für die Schriftsprache der Gegenwart: Arabisch-Deutsch. Edited by: Lorenz Kropfitsch. Otto Harrassowitz Verlag.

السيرة الذاتية

المُعتزّ بالله السَّعيد
يَعمل مُدَرِّسًا بقسم علم اللُّغة والدِّراسات السَّامِيَّة والشَّرقِيَّة في كُلِّيَّة دار العُلُوم بجامعةالقاهرة، وباحثًا في حوسبة اللُّغة والمُعجم بالمركز العربيّ للأبحاث ودراسة السِّياسات في الدَّوحة.حصلَ من جامعة القاهِرة على درجة الدُّكتوراهفي علم اللُّغة والدِّراسات السَّاميَّة والشَّرقِيَّة. نَشَرَ ما يقرُب من عشرين ورقة بحثِيَّة في دوريَّاتٍ علميَّةٍ ومؤتمراتٍ دوليَّة، بالإضافة إلى خمسة كُتُبٍ في المُعجَمِيَّة العَرَبِيَّة والدِّراسات اللُّغَوِيَّة المُعاصِرة؛ منها: علم اللُّغة والتِّقنيات المُعاصرة، علم الدِّلالة ونظريَّة المعنى، مُقَدِّمة في حَوسَبة اللُّغة العربيَّة "بالاشتراك"؛ كما ساهمَ في أكثر من عشرة مشرُوعاتٍ بحثِيَّةٍ دوليَّةٍ تُعنى بحوسَبة اللُّغة العربيَّة وتقنياتها، منها: مشرُوع مُعجم الدَّوحة التَّاريخيّ للُّغة العربيَّة، ومشرُوع بناء شبكة دلاليَّة للُّغة العربيَّةSEWAC، ومشرُوع بناء نظام الخليل للتَّحليل الصَّرفيّ للُّغة العربيَّة. شاركَ في الإشراف على عددٍ من الأطروحات العلميَّة في مصرَ وإسبانيا؛ وهو عُضوٌ بهيئة تحرير المجلَّة الدَّوليَّة للتَّقدُّم والبحث العلميّIJSPR والمجلَّة الدَّوليَّة لعُلُوم وهندسة الحاسب "باللُّغة العَرَبيَّة" IJCSEA وعُضوٌ باللِّجان العلميَّة لعدد من المُؤتمرات الدَّوليَّة، منها: المُؤتمر الدَّوليّ لتطبيقات اللُّغات الطَّبيعيَّة CEC-TAL 2015، 2015، سوسة، تُونُس؛ والنَّدوة الدَّوليَّة الخامسة حول المعالَجة الآليَّة للُّغة العربيَّة CITALA'14، 2014م، وجدة، المغرب؛ وورشة عمل المعالجة الآليَّة للُّغة العربيَّة EMNLP 2014، 2014م، الدَّوحة، قَطَر. حصلَ في عام 2012 على جائزة المُنَظَّمة العربيَّة للتَّربية والثَّقافة والعُلُوم (ألكسو ALECSO) للإبداع والابتكار التِّقَنِيّ [المركز الأوَّل] في مَيدان "المَعلُوماتيَّة والمُعالَجة الآليَّة للُّغة العربيَّة".

Towards an Arabic WordNet for Lexicography Purposes

Almoataz B. Al-Said
Faculty of Dar Al Uloom, Cairo University, Egypt
moataz@cu.edu.eg

Abstract: ***This study seeks to provide a methodology for building and developing an Arabic WordNet for lexicography purposes. The study aims to set a clear vision of the mechanisms for building an Arabic WordNet, taking into account the nature of the Arabic language and its lexical system; so that we can employ these mechanisms in the future in various forms of Arabic lexicography.The study comes in six main sections including the introduction then a presentation of the main problems, followed by an introduction of the proposed methodology to build the Arabic WordNet and how can we employ the WordNet in Arabic lexicography. Finally, a review of the results is introduced then the conclusions are discussed.***

Keywords:
WordNet, Ontology, Lexicography, Semantic Relations