ما هي العربية؟

What is "Arabic"?

في سياق هذا الكتاب، ستُستخدم لفظة "العربية/العربي" للإشارة إلى اللغة، ومجموعة اللهجات وأيضاً الخط.

١,١ اللغة العربية واللهجات العربية

تتكون اللغة العربية من مجموعة من متغيرات لغوية متعددة، من بين هذه المتغيرات اللغوية متغير معين له حالة خاصة باعتباره المعيار الكتابي الرسمي للإعلام، والثقافة، والتعليم في جميع أنحاء العالم العربي[1]. أما بالنسبة للمتغيرات الأخرى فهي اللهجات المنطوقة غير الرسمية التي تستخدم وسيلة للتواصل في الحياة اليومية. وبطبيعة الحال، فاللغة موجودة في سلسلة متصلة، تاريخياً وجغرافياً على حد سواء. إن مصطلح اللغة مقابل اللهجة ليس سوى تعبير عن نفوذ وهيمنة مجموعة أو إيديولوجية

(١) يشير العالم العربي إلى البلدان الناطقة بالعربية التي تمتد بين المحيط الأطلسي والخليج العربي. ويصل عدد سكان هذه الدول مجتمعة لأكثر من ٣٠٠ مليون نسمة. الدول الاثنان والعشرون الأعضاء في جامعة الدول العربية هي: الجزائر، البحرين، جزر القمر، جيبوتي، مصر، العراق، الأردن، الكويت، لبنان، ليبيا، موريتانيا، المغرب، عمان، فلسطين، قطر، السعودية، الصومال، السودان، سوريا، تونس، الإمارات العربية المتحدة، اليمن. كما أن اللغة العربية تعتبر لغة رسمية في ثلاث دول أخرى وهي تشاد وأريتريا وإسرائيل. وتعتبر إيران وباكستان وأفغانستان وتركيا بلداناً مسلمة مجاورة وقريبة من العالم العربي، لكنها لا تتحدث العربية.

obeikandl.com

على الأخرى. وفي العالم العربي، السياسة (ويقصد بها القومية العربية) والدين (ويقصد به الإسلام) هما ما يشكل الرؤية حول ما هي اللغة العربية، وما هي اللهجة العربية. علاقة النفوذ هذه مشابهة لغيرها من العلاقات التي توجد بين اللغات ولهجاتها. ومع ذلك، فإن هناك جانبين مميزين للوضع اللغوي في اللغة العربية، هما: (أ) الدرجة العالية من الاختلاف بين اللغة العربية الفصحى ولهجاتها و(ب) حقيقة أن العربية الفصحى ليست اللغة الأم لأي عربي[١].

تعتبر اللغة العربية المعاصرة/فصحى العصر (Modern Standard Arabic (MSA)) [١] اللغة الرسمية في العالم العربي. وهي اللغة الأساسية في وسائل الإعلام والتعليم. وتعتمد اللغة العربية المعاصرة في نحوها وصرفها وصوتيتها على العربية الكلاسيكية اللغة العربية الفصحى/فصحى التراث (Classical Arabic) [1]، التي هي لغة القرآن الكريم. أما لفظياً، فاللغة العربية المعاصرة أكثر حداثة، وتستخدم في الكتابة بشكل أساسي وليس في التخاطب.

في المقابل، فإن اللهجات العربية هي الصورة الواضحة للغة الأم وينحصر استخدامها في التواصل اليومي غير الرسمي. لا يتم تدريس هذه اللهجات في المدارس أو حتى توحيدها على الرغم من استخدامها في القصص الشعبية والأغاني والأفلام والبرامج التلفزيونية، فهي في الأساس منطوقة وغير مكتوبة. ومع ذلك، بدأ هذا الوضع في التغير مع قدرة الكثير من العرب على التواصل عبر وسائل الإعلام الإلكترونية مثل رسائل البريد الإلكتروني والمجموعات الإخبارية. من ناحية تاريخية نجد أن اللهجات العربية غير مرتبطة ارتباطاً تاماً باللغة العربية الفصحى، فهي نتيجة

(١) قارنها مثلا مع اللغة الألمانية العليا (High German)، التي تعتبر لهجة حية بقواعد قياسية.

openkandil.com

للتفاعل بين لهجات مختلفة من اللغة العربية الفصحى[١] واللغات الأخرى الموجودة في دول الجوار أو اللغات الموجودة سابقاً فيما يسمى اليوم بالعالم العربي. على سبيل المثال، اللغة العربية المستخدمة في الجزائر متأثرة باللغة البربرية وكذلك الفرنسية. فاللهجات العربية يختلف بعضها كثيراً عن بعض وعن اللغة العربية الفصحى من الناحية الصوتية والصرفية والنحوية والاختيار اللفظي.

تختلف اللهجات العربية عن بعضها من عدة أبعاد من أهمها البعد الجغرافي والاجتماعي. جغرافياً، يمكن تقسيم العالم العربي لغوياً بعدة طرق مختلفة. فيما يلي سرد لواحدة من هذه التقسيمات العديدة (ولا ينبغي أن يؤخذ هذا التقسيم على أنه يعني أن جميع أعضاء مجموعة اللهجات متجانسة تماماً لغوياً):

- اللهجة المصرية (Egyptian Arabic) تغطي لهجات أهل النيل: مصر والسودان.
- اللهجة الشامية (Levantine) وتشمل اللهجات في لبنان، وسوريا، والأردن، وفلسطين، وإسرائيل.
- اللهجة الخليجية (Gulf Arabic) تتضمن لهجات أهل الكويت، والإمارات العربية المتحدة، والبحرين، وقطر. وعادة تضمن المملكة العربية السعودية على الرغم من أن هناك مجموعة واسعة من اللهجات الفرعية داخلها. كما يتم تضمين اللهجة العمانية في بعض الأحيان.
- لهجة شمال أفريقيا (Maghrebi) وتغطي لهجات المغرب، والجزائر، وتونس، وموريتانيا. وتشمل في بعض الأحيان اللهجة الليبية.
- اللهجة العراقية (Iraqi Arabic) وهي مزيج ما بين اللهجة الشامية والخليجية.

---

(١) لهجات القبائل العربية في شبه الجزيرة (توضيح من المترجمة).

- اللهجة اليمنية العربية (Yemenite Arabic) في الغالب لها فئتها الخاصة.
- اللهجة المالطية (Maltese Arabic) في الغالب لا تعتبر لهجة عربية، إنما هي نوع من أنواع اللغة العربية التي تعتبر لغة مستقلة بذاتها تكتب بالحروف اللاتينية.

من الشائع اجتماعياً التمييز بين لهجات ثلاث فرعية لكل لهجة داخل المنطقة وهي: المدينة والريف والبدو. وترتبط درجات اللهجات الفرعية الثلاث في كثير من الأحيان بالتسلسل الهرمي للهجات، فالطبقة الغنية من سكان المدن هم على رأس الهرم بينما البدو يمثلون القاع. كما أن مختلف الارتباطات الاجتماعية متواجدة كما هو الحال في العديد من اللغات الأخرى في جميع أنحاء العالم. على سبيل المثال، تعتبر لهجة أهل المدينة أقل معلمة (بفتح اللام وشدها) وأكثر دقة ومرموقة، في حين تعتبر اللهجة البدوية أقل مستوى وأكثر صرامة، وفي ذات الوقت أقرب إلى أصل اللغة. ومن المعروف أن المتحدثين ينتقلون باعتدال بين اللهجات بناء على السياقات الاجتماعية المختلفة [2].

العلاقة بين اللغة العربية المعاصرة واللهجة في منطقة معينة معقدة نوعاً ما؛ فالعرب لا يفكرون بهما كلغات منفصلة. هذا التصور يؤدي إلى نوع خاص من التعايش بين شكلين من اللغة يخدمان أغراضاً مختلفة. هذه الحالة هي ما يطلق عليه اللغويون مصطلح الازدواج اللغوي diglossia [3] . وعلى الرغم من أن النوعين لهما مجالان من الاستخدام: الكتابة الرسمية (اللغة العربية المعاصرة) مقابل التخاطب غير الرسمي (اللهجة)، إلا أن هناك مساحة متداخلة كبيرة بينهما تُملأ في كثير من الأحيان بمزيج من هذين الشكلين [1 ، 2].

١,٢ الخط العربي

يستخدم الخط العربي لكتابة اللغة العربية، كما يستخدم أيضا لكتابة العديد من اللغات في جميع أنحاء العالم التي لا علاقة لها باللغة العربية مثل الفارسية، والكردية، والأوردية، والباشتوية[١]. في الواقع، بعض هذه اللغات هي أقرب إلى اللغة الإنجليزية منها إلى اللغة العربية: على سبيل المثال، الفارسية هي لغة من العائلة الهند- أوروبية (Indo-European)، وهي تضم أيضاً الإنجليزية والفرنسية.

تكتب اللهجات العربية عادة باستخدام الخط العربي على الرغم من عدم وجود معيار نظامي لطريقة إملاء اللهجات. وفي خلال القرن الماضي، كانت هناك مطالبات متفرقة لاستبدال الخط العربي بالخط اللاتيني للغة العربية المعاصرة أو على الأقل للهجاتها. تزامنت هذه المطالبات مع نجاح برنامج كمال أتاتورك للرومنة في تركيا، حيث كان الخط العربي مستخدماً في كتابة اللغة التركية خلال الحكم العثماني. وقد حافظت المعارضة السياسية والدينية لمثل هذه المطالبات على استخدام الخط العربي في العالم العربي. حالياً، ينظر في النداءات التي تدعو لتغيير إملاء اللهجات في الخط العربي على أنها تحدٍ لسلطة ونفوذ اللغة العربية المعاصرة.

١,٣ محتوى الكتاب

يتكون هذا الكتاب من ثمانية فصول بما في ذلك هذا الفصل التمهيدي. وعلى الرغم من تعدد صور العربية، سيكون تركيزنا في هذا الكتاب في المقام الأول على اللغة العربية المعاصرة، وقد نشير أحياناً إلى اللغة العربية الفصحى أو اللهجات العربية. كما سيركز الكتاب على اللغة ذاتها بدلاً من التركيز على الطرق البحثية لمعالجة اللغة الطبيعية، وذلك خوفاً من الخوض في مشاكل معالجة اللغة الطبيعية.

---

(١) اللغة الوطنية في أفغانستان (توضيح من المترجمة).

obeikandl.com

يحتوي كل فصل على عنصر لغوي يتبعه نقاشات حول طرق معالجته حاسوبياً، بحيث يتم تعريف مهام معالجة اللغة الطبيعية المتبعة وأمثلة عليها، وأيضاً توفير إشارات مرجعية إلى بحوث سابقة وجارية في هذا المجال.

وبما أن الكتاب يعتبر مقدمة في هذا المجال الواسع، فإنه يسعى لتوفير نقطة انطلاق للباحثين للحصول على مزيد من المعلومات حول الوظائف المختلفة لمعالجة اللغات الطبيعية. وقد صمم الكتاب للحاسوبيين واللغويين على حد سواء ولا حاجة لمعرفة مسبقة باللغة العربية[١].

في الفصل الثاني، نقدم الخط العربي وخصائصه. ونناقش أيضاً قضايا تنوع الترميز، والنقل الكتابي/النقحرة (transliteration) والتعرف الضوئي على الحروف (optical character recognition). كما سنشير بإيجاز لقضايا ذات صلة بلغات أخرى غير العربية تستخدم الخط العربي. وسنناقش المهارات الأساسية للتعامل مع النص العربي، حتى لمن يجهلون الخط العربي.

في الفصل الثالث، سنناقش علم الأصوات العربية والتهجئة (كيفية استخدام الخط العربي لتمثيل اللغة العربية). وسنناقش أيضا مشاكل معالجة اللغة العربية مع أسماء العلم المنقولة كتابياً/المنقحرة (proper name transliteration)، والتشكيل (diacritization)، والتعرف وتوليد الكلام آلياً (automatic speech recognition and speech synthesis).

في الفصل الرابع، سنناقش الثراء الصرفي للغة العربية، حيث سيركز الفصل على المصطلحات في هذا المجال ويقدم شرحاً مقتضباً للصرف العربي.

---

(١) حيث إن الكتاب موجه للقارئ غير العربي في الأصل.

في الفصل الخامس، سنتناول عمليات التحليل والتوليد الصرفي للغة العربية ، وكشف اللبس الصرفي (morphological disambiguation) والتقطيع (tokenization) وتعيين أقسام الكلام (POS tagging).

في الفصل السادس، سنتكلم عن بناء الجملة العربية. وسنستعرض باختصار القضايا النحوية للغة العربية، ونناقش الموارد البحثية المتاحة والمخصصة لتحليل وبناء اللغة العربية.

في الفصل السابع، سنتحدث بإيجاز عن الدلالات العربية حاسوبياً، وعلى وجه التحديد وصف بعض الموارد المتاحة لإجراء البحوث في هذا المجال.

وفي الفصل الثامن سنتحدث باختصار عن الترجمة الآلية للغة العربية.

ومن الجدير ذكره، أن أحجام الفصول تتوافق بشكل أو بآخر مع ما يميز اللغة العربية لغوياً، مع التركيز على صرف اللغة العربية، يليه الخط العربي. وينتهي الكتاب بأربعة ملاحق تحتوي على إشارات مرجعية لموارد وأدوات ذات صلة بمعالجة اللغة العربية.

وكملاحظة أخيرة بالنسبة للترجمة والكتابة الصوتية للغة العربية، فقد استخُدام مخطط (نظام نسخ) حبش – سودي- بكوالتر [3] (Habash-Soudi-Buckwalter) لتمثيل التهجئة العربية وأصواتها. وسنناقش هذا المخطط في الفصل الثاني من الكتاب.

obeikandl.com