WIBARAB Feature Database Model

Daniel Schopper

# WIBARAB Feature Database Model

## Version 0.1

|  |  |
| --- | --- |
| Author(s): | Daniel Schopper |
| Version: | 0.1 |
| Last Changed: | 2022-05-17 |
| Documentation generated: | 2022-05-20T12:32:22.584902+02:00 |

## Graphical Overview

![](data:image/svg+xml;base64;base64,)

Graphical Overview of the data model.  
Dashed lines mean implicit shortcut relations.

## Description

This document describes the abstract model of the WIBARAB Feature Database.

The WIBARAB Database of Lingustic Features systematically collects linguistic features found in Arabic varieties and categorizes how they are realised in the varieties in question. The database draws its content from published resources gathered in the [VICAV library](https://www.zotero.org/groups/2165756/vicav) (e.g. language atlases or grammar books) as well as material collected during fieldwork in the WIBARAB project, structuring and contextualising it.

The main entities in the database are so-called [<FEATURE REALISATIONS>](#FeatureRealisation): For each feature and each variety, evidence is collected from the above-mentioned sources how a given feature is realised. Since the aim of the database is to facilitate cross-variety comparison, the single realisations are grouped into categories of distinct FEATURE VALUES so relations between dialects can be interpolated by juxtaposing their (dis)agreeing treatment of such features.

The resulting database will offer the following basic functionality to the users:

* Plot groups of FEATURE REALISATIONS on a map, indicating their FEATURE VALUE by different symbols and/or colours. The locations of the points are determined by a fictive geolocation representing the variety.
* The user should be able to filter the map points by the temporal dimension of the source data (more specifically, the period in which the data of the underlying publication has been collected).
* ... ?

## Classes

### Class: "Feature"

#### Definition

A linguistic phenomenon of freely definable dimension / size. Following [WALS’ definition](https://wals.info/feature): "A feature is a structural property of language that describes one aspect of cross-linguistic diversity." Each feature can be realised in a finite number of ways, each of which is called a [<FEATURE VALUE>](#FeatureTypes).

#### Properties

| Property Name / Cardinality | Data Type | Remarks |
| --- | --- | --- |
| Cardinality: 1 | [TEI Markup](#teiContent) |  |
| **Name** Cardinality: 1 | [short text](#shortText) |  |

#### Relations (outgoing)

* [<Feature>](#Feature) [has possible realisations (is possible realisation of)](#FeatureHasPossibleRealisations) [<Feature Value>](#FeatureTypes)

#### Relations (incoming)

* [<Feature Realisation>](#FeatureRealisation) [is an instance of (has instances)](#X1107ed17e8a1fbc28a28107732dead00e3068a3) [<Feature>](#Feature)

### Class: "Feature Value"

#### Definition

The abstract notion of how a [<FEATURE>](#Feature) can be realised.

NOTE This is what WALS calls "feature values".

TODO add examples

who: AD when: 2022-05-17 status: new

#### Properties

| Property Name / Cardinality | Data Type | Remarks |
| --- | --- | --- |
| **Description** Cardinality: 0-1 | [TEI Markup](#teiContent) | A description/discussion of the feature value in question. |
| **Identifier** Cardinality: 1 | [Identifier](#identifier) |  |
| **Name** Cardinality: 1 | [short text](#shortText) | a short label representing the feature value |

#### Relations (outgoing)

No outgoing relations defined.

#### Relations (incoming)

* [<Feature>](#Feature) [has possible realisations (is possible realisation of)](#FeatureHasPossibleRealisations) [<Feature Value>](#FeatureTypes)
* [<Feature Realisation>](#FeatureRealisation) [is of type (has instances)](#featureRealisationIsOfType) [<Feature Value>](#FeatureTypes)

### Class: "Feature Realisation"

#### Definition

Represents the assertion found in some [<SOURCE>](#Source) that a [<DIALECT>](#Dialect) [*exhibits*](#DialectExhibitsFeatureRealisation) [<FEATURE REALISATION>](#FeatureRealisation) in a specific way (expressed by [<FEATURE REALISATION>](#FeatureRealisation) [*is of type*](#featureRealisationIsOfType) [<FEATURE VALUE>](#FeatureTypes)).

NOTE

Any constraints

#### Properties

| Property Name / Cardinality | Data Type | Remarks |
| --- | --- | --- |
| **Note** Cardinality: 0-n | [long text](#longText) |  |

#### Relations (outgoing)

* [<Feature Realisation>](#FeatureRealisation) [asserted by (asserts)](#featureRealisationHasSource) [<Source>](#Source)
* [<Feature Realisation>](#FeatureRealisation) [has constraint (is constraining)](#featureRealisationHasConstraint) [<Realisation Constraint>](#Constraint)
* [<Feature Realisation>](#FeatureRealisation) [is an instance of (has instances)](#X1107ed17e8a1fbc28a28107732dead00e3068a3) [<Feature>](#Feature)
* [<Feature Realisation>](#FeatureRealisation) [is of type (has instances)](#featureRealisationIsOfType) [<Feature Value>](#FeatureTypes)

#### Relations (incoming)

* [<Dialect>](#Dialect) [exhibits (is exhibited in)](#DialectExhibitsFeatureRealisation) [<Feature Realisation>](#FeatureRealisation)

### Class: "Dialect"

#### Definition

A distinct language variety spoken at a given location by a given group.

CHECKME is this accurate or should it be renamed to "variety", "vernacular" … ?

who: DS when: 2022-05-17 status: new

#### Properties

| Property Name / Cardinality | Data Type | Remarks |
| --- | --- | --- |
| **ID** Cardinality: 1 | [Identifier](#identifier) | the dialect identifiers are made up of a three-letter country code + hyphen + three-letter settlement code |
| **Name** Cardinality: 1 | [short text](#shortText) |  |
| **type of community** Cardinality: 0-n | [Vocabulary Term](#vocabTerm) in [community types](#vocabs.communityTypes) |  |

#### Mapping to Guidelines of the Text Encoding Initiative. Version P5

![](data:image/svg+xml;base64;base64,)

Listing 1

#### Relations (outgoing)

* [<Dialect>](#Dialect) [exhibits (is exhibited in)](#DialectExhibitsFeatureRealisation) [<Feature Realisation>](#FeatureRealisation)
* [<Dialect>](#Dialect) [has person in charge (is in charge of)](#dialectHasPersonInCharge) [<Person>](#Person)
* [<Dialect>](#Dialect) [is spoken at (is place where is spoken)](#dialectIsSpokenAtLocation) [<Place>](#Place)

#### Relations (incoming)

* [<Source>](#Source) [provides information on (is described in)](#sourceProvidesInformationOn) [<Dialect>](#Dialect)

### Class: "Source"

**Superclass of:** [<BIBLIGRAPHIC ITEM>](#BibliographicItem), [<CAMPAIGN>](#Campaign), [<PERSONAL COMMUNICATION>](#PersonalCommunication)

#### Definition

Any Source

#### Properties

| Property Name / Cardinality | Data Type | Remarks |
| --- | --- | --- |
| **period of collection** Cardinality: 1 | [period](#period) | the date when the underlying information of this source was collected. In case of publication which do not make this information transparent, the decade before the year of publication is assumed as a default value. |

#### Relations (outgoing)

* [<Source>](#Source) [provides information on (is described in)](#sourceProvidesInformationOn) [<Dialect>](#Dialect)

#### Relations (incoming)

* [<Feature Realisation>](#FeatureRealisation) [asserted by (asserts)](#featureRealisationHasSource) [<Source>](#Source)
* [<Original Source>](#OriginalSource) [is taken from (is quoted in)](#exampleTakenFrom) [<Source>](#Source)

### Class: "Bibligraphic Item"

**Subclass of:** [<SOURCE>](#Source)

#### Definition

A bibliographic item (held in the VICAV Zotero Group Library) which is the source of some data in the database.

NOTE We don't want to list all fields of a full-blown bibliographic record here …

#### Properties

| Property Name / Cardinality | Data Type | Remarks |
| --- | --- | --- |
| **ID** Cardinality: 1 | [Identifier](#identifier) |  |
| **period of collection** Cardinality: 1 | [period](#period) | the date when the underlying information of this source was collected. In case of publication which do not make this information transparent, the decade before the year of publication is assumed as a default value.  Inherited from [<SOURCE>](#Source) |

#### Relations (outgoing)

No outgoing relations defined.

#### Relations (incoming)

* [<Original Source>](#OriginalSource) [is taken from (is quoted in)](#originalSourceTakenFrom) [<Bibligraphic Item>](#BibliographicItem)

### Class: "Campaign"

**Subclass of:** [<SOURCE>](#Source)

#### Definition

A field campaign during WIBARAB collecting evidence on feature realisations.

#### Relations (outgoing)

No outgoing relations defined.

#### Relations (incoming)

No outgoing relations defined.

### Class: "Personal Communication"

**Subclass of:** [<SOURCE>](#Source)

#### Definition

Some type of personal communication (e-mail exchange, ...) which is treated as evidence for a feature realisation assertion.

#### Relations (outgoing)

No outgoing relations defined.

#### Relations (incoming)

No outgoing relations defined.

### Class: "Place"

#### Definition

A place of any type/extent from the [VICAV place list](https://github.com/acdh-oeaw/vicav-content/blob/b0feba5e3a3799aaecbf259c0b6da47bc9b8d674/vicav_biblio/vicav_geodata.xml).

#### Properties

| Property Name / Cardinality | Data Type | Remarks |
| --- | --- | --- |
| **Geonames ID** Cardinality: 0-1 | [Identifier](#identifier) |  |

#### Relations (outgoing)

No outgoing relations defined.

#### Relations (incoming)

* [<Dialect>](#Dialect) [is spoken at (is place where is spoken)](#dialectIsSpokenAtLocation) [<Place>](#Place)

### Class: "Person"

#### Definition

Any person in the database, including project members or informants.

#### Relations (outgoing)

* [<Person>](#Person) [is informant in (has informants)](#personIsInformantOfSource) [<Campaign>](#Campaign), [<Personal Communication>](#PersonalCommunication)

#### Relations (incoming)

* [<Campaign>](#Campaign), [<Personal Communication>](#PersonalCommunication) [collected by (collected)](#sourceHasBeenCollectedByPerson) [<Person>](#Person)
* [<Dialect>](#Dialect) [has person in charge (is in charge of)](#dialectHasPersonInCharge) [<Person>](#Person)

### Class: "Speaker Group"

#### Definition

A group of speakers, defined by some common property.

#### Properties

| Property Name / Cardinality | Data Type | Remarks |
| --- | --- | --- |
| **Group Type** Cardinality: 1-n | [Vocabulary Term](#vocabTerm) in [Group Types Vocabulary](#vocabs.groupTypes) |  |

#### Relations (outgoing)

No outgoing relations defined.

#### Relations (incoming)

No outgoing relations defined.

### Class: "Realisation Constraint"

#### Definition

Represents that a [<DIALECT>](#Dialect) [*exhibits*](#DialectExhibitsFeatureRealisation) [<FEATURE REALISATION>](#FeatureRealisation) only in certain contexts.

CHECKME All kinds of social-linguistic constraints could also be modelled as "speaker groups" with the broadest group being all speakers of a dialect, indiacting "No constraint".

who: DS when: 2022-05-17 status: new

#### Properties

| Property Name / Cardinality | Data Type | Remarks |
| --- | --- | --- |
| **Constraint Type** Cardinality: 1-n | [Vocabulary Term](#vocabTerm) in [Constraint Types](#vocabs.constraintTypes) |  |
| **academic level** Cardinality: 0-n | [Vocabulary Term](#vocabTerm) in [academic level constraint values](#vocabs.academicLevels) | indicates that the feature realisation is salient in a group of speakers with a common level of education |
| **age group** Cardinality: 0-n | [Vocabulary Term](#vocabTerm) in [age group constraint values](#vocabs.ageGroups) |  |
| **gender constraint** Cardinality: 0-n | [Vocabulary Term](#vocabTerm) in [gender constraint values](#vocabs.gender) |  |
| **lexical domain** Cardinality: 0-1 | [short text](#shortText) | indicates that the feature realisation is salient in a group of speakers with a common professional background  CHECKME Given that it might be difficult to come up with a closed vocabulary of occupational groups, this is defined as a short text.  who: DS when: 2022-05-02 status: new |
| **occupational group** Cardinality: 0-1 | [short text](#shortText) | indicates that the feature realisation is salient in a group of speakers with a common professional background  CHECKME CHECKME: Given that it might be difficulat to come up with a closed vocabulary of occupational groups, this is defined as a short text.  when: 2022-05-02 status: new who: DS |

#### Relations (outgoing)

No outgoing relations defined.

#### Relations (incoming)

* [<Feature Realisation>](#FeatureRealisation) [has constraint (is constraining)](#featureRealisationHasConstraint) [<Realisation Constraint>](#Constraint)

### Class: "Full Text"

**Superclass of:** [<ORIGINAL SOURCE>](#OriginalSource), [<EXAMPLE>](#Example)

#### Definition

Any kind of (annotated) text.

#### Properties

| Property Name / Cardinality | Data Type | Remarks |
| --- | --- | --- |
| **Quotation** Cardinality: 1 | [TEI Markup](#teiContent) |  |

#### Relations (outgoing)

No outgoing relations defined.

#### Relations (incoming)

No outgoing relations defined.

### Class: "Original Source"

**Subclass of:** [<FULL TEXT>](#FullText)

#### Definition

A quotation taken from the bibliographic record (not an example of the feature in question)

#### Mapping to Guidelines of the Text Encoding Initiative. Version P5

tei:cit[@type = "originalSource"]

#### Relations (outgoing)

* [<Original Source>](#OriginalSource) [is taken from (is quoted in)](#originalSourceTakenFrom) [<Bibligraphic Item>](#BibliographicItem)
* [<Original Source>](#OriginalSource) [is taken from (is quoted in)](#exampleTakenFrom) [<Source>](#Source)

#### Relations (incoming)

No outgoing relations defined.

### Class: "Example"

**Subclass of:** [<FULL TEXT>](#FullText)

#### Definition

An (annotated) example for the [<FEATURE TYPE>](#FeatureTypes) in question.

#### Mapping to Guidelines of the Text Encoding Initiative. Version P5

tei:cit[@type = "originalSource"]

#### Relations (outgoing)

No outgoing relations defined.

#### Relations (incoming)

No outgoing relations defined.

## Relations

### asserted by

asserts

**Cardinality:** 1-n / 1  
**Source Class:** [<FEATURE REALISATION>](#FeatureRealisation)  
**Target Class:** [<SOURCE>](#Source)

##### has possible realisations

is possible realisation of

**Cardinality:** 1 / 1-n  
**Source Class:** [<FEATURE>](#Feature)  
**Target Class:** [<FEATURE VALUE>](#FeatureTypes)

##### is an instance of\*

has instances

**Cardinality:** 1 / 1  
**Source Class:** [<FEATURE REALISATION>](#FeatureRealisation)  
**Target Class:** [<FEATURE>](#Feature)

##### is of type

has instances

**Cardinality:** 1 / 1-n  
**Source Class:** [<FEATURE REALISATION>](#FeatureRealisation)  
**Target Class:** [<FEATURE VALUE>](#FeatureTypes)

##### has constraint

is constraining

**Cardinality:** 1 / 0-n  
**Source Class:** [<FEATURE REALISATION>](#FeatureRealisation)  
**Target Class:** [<REALISATION CONSTRAINT>](#Constraint)

##### is spoken at

is place where is spoken

**Cardinality:** 1-n / 1-n  
**Source Class:** [<DIALECT>](#Dialect)  
**Target Class:** [<PLACE>](#Place)

NOTE Documents where a dialect is typically spoken in a place.

CHECKME check if n-to-n relation is ok

who: DS when: 2022-05-17 status: new

##### has person in charge

is in charge of

**Cardinality:** 1-n / 1  
**Source Class:** [<DIALECT>](#Dialect)  
**Target Class:** [<PERSON>](#Person)

NOTE internal procedural information who is the main person in charge of a dialect

##### exhibits

is exhibited in

**Cardinality:** 1 / 1-n  
**Source Class:** [<DIALECT>](#Dialect)  
**Target Class:** [<FEATURE REALISATION>](#FeatureRealisation)

##### provides information on

is described in

**Cardinality:** 1 / 1-n  
**Source Class:** [<SOURCE>](#Source)  
**Target Class:** [<DIALECT>](#Dialect)

##### collected by

collected

**Cardinality:** 1 / 1-n  
**Source Class:** [<CAMPAIGN>](#Campaign)[<PERSONAL COMMUNICATION>](#PersonalCommunication)  
**Target Class:** [<PERSON>](#Person)

##### is informant in

has informants

**Cardinality:** 1-n / 1-n  
**Source Class:** [<PERSON>](#Person)  
**Target Class:** [<CAMPAIGN>](#Campaign)[<PERSONAL COMMUNICATION>](#PersonalCommunication)

##### is taken from

is quoted in

**Cardinality:** 1-n / 1  
**Source Class:** [<ORIGINAL SOURCE>](#OriginalSource)  
**Target Class:** [<BIBLIGRAPHIC ITEM>](#BibliographicItem)

##### is taken from

is quoted in

**Cardinality:** 1-n / 1  
**Source Class:** [<ORIGINAL SOURCE>](#OriginalSource)  
**Target Class:** [<SOURCE>](#Source)

## Vocabularies

| Vocabulary Name | Example Values | Remarks |
| --- | --- | --- |
| gender constraint values | dominant feminine usage / dominant masculine usage | defines a set of values describing realisation constraints for the gender dimension  (DS: this seems a somewhat problematic dichotomy) |
| academic level constraint values |  | TODO define! |
| age group constraint values | children / adolescents / young adults / adults / middle aged / elders / very old | defines a set of values describing realisation constraints for the speaker age dimension  (DS: this is a placeholder list and should be replaced by something more sensible) |
| community types | tribe |  |
| Constraint Types | Gender / Lexical / Period / Age Group / Academic |  |
| Group Types Vocabulary | Tribe |  |

## Datatypes

| Name | Used by | Remarks |
| --- | --- | --- |
| Identifier | [BibliographicItem.ID](#bibliographicItem.ID), [Dialect.ID](#dialect.ID), [FeatureTypes.Identifier](#featureValue.ID), [Place.Geonames ID](#place.geonameID) | a identifier unique in the scope of the project |
| TEI Markup | [FeatureTypes.Description](#featureValue.description), [Feature.](#feature.discussion), [FullText.Quotation](#fulltext.quotation) | XML fragment conforming to the [TEI Guidelines](https://tei-c.org/guidelines/), potentially including mixed content |
| Vocabulary Term | [Dialect.type of community](#dialect.typeOfCommunity), [Constraint.Constraint Type](#constraint.type), [Constraint.academic level](#X7d646ff5f88ef8f149c999f4f813bf92561a8a7), [Constraint.age group](#featureRealisation.ageGroupConstraint), [Constraint.gender constraint](#featureRealisation.genderConstraint), [Group.Group Type](#group.type) | a term from a controlled vocabulary |
| long text | [FeatureRealisation.Note](#featureRealisation.note) | a long text (more than 256 characters long |
| period | [BibliographicItem.period of collection](#bibliographicitem.source.collectionDate), [Source.period of collection](#source.collectionDate) | some range of dates (e.g. decades) |
| short text | [Dialect.Name](#dialect.name), [FeatureTypes.Name](#featureValue.name), [Feature.Name](#feature.name), [Constraint.lexical domain](#Xa67ab80eca7c81e2eb6c3a89b70bae412eaa99b), [Constraint.occupational group](#Xb5ca1766ae927c4d3290f7cc74a83c2933fa12a) | a short text (less or equal to 256 characters long) |

## References

Guidelines of the Text Encoding Initiative. Version P5TEI Guidelines Version 4.4.0. Last updated on 19th April 2022, revision ff9cc28b0. https://www.tei-c.org/release/doc/tei-p5-doc/en/html/