Paralex specifications (default package)

This is a human-readable rendition of a JSON file defining a frictionless package. It was generated automatically.

- `name`: paralex
- `licenses`:
  - GPL-3.0: GNU General Public License 3.0
- `keywords`: lexicon, inflection, linguistics, morphology, paradigms
-
- `profile` data-package
- `contributors`
  - [1]
    - `* title` Sacha Beniamine
    - `* role` maintainer
  - [2]
    - `* title` Cormac Anderson
    - `* role` contributor
  - [3]
    - `* title` Jules Bouton
    - `* role` contributor
  - [4]
    - `* title` Mae Carroll
    - `* role` contributor
  - [5]
    - `* title` Borja Herce
    - `* role` contributor
  - [6]
    - `* title` Matías Guzmán Naranjo
    - `* role` contributor
  - [7]
    - `* title` Matteo Pellegrini
    - `* role` contributor
  - [8]
    - `* title` Erich Round
    - `* role` contributor
  - [9]
    - `* title` Helen Sims-Williams
    - `* role` contributor
- `version` 2.0.8
- `languages_iso639` ['']
- `related_identifiers`
  - [1]
    - `* identifier`
    - `* relation` <Any of isCitedBy, cites, isSupplementTo, isSupplementedBy, isContinuedBy, continues, isDescribedBy, describes,

hasMetadata, isMetadataFor, isNewVersionOf, isPreviousVersionOf, isPartOf, hasPart, isReferencedBy, references, isDocumentedBy, documents, isCompiledBy, compiles, isVariantFormOf, isOriginalFormof, isIdenticalTo, isAlternateIdentifier, isReviewedBy, reviews, isDerivedFrom, isSourceOf, requires, isRequiredBy, isObsoletedBy, obsoletes>

This package describes the following tables:

forms

Inflected forms - This file is located in `forms.csv`.

- The identifier column (or `primaryKey`) is `['form_id']`
- **Formal relations (foreignKeys)** with other tables:

Each value in column	Must refer to
`['cell']`	`['cell_id']` in table `cells`
`['lexeme']`	`['lexeme_id']` in table `lexemes`

Columns defined by `forms-schema`:

- `form_id` (`string`): Form table row identifiers. These identifiers are specific to form, lexeme, cell triples.
    - constraints: a `form_id` is obligatory; it must be unique.
- `lexeme` (`string`): Reference to a lexeme identifier. Lexeme identifiers must be unique to paradigms.
    - constraints: a `lexeme` is obligatory.
    - `rdfProperty`: https://www.paralex-standard.org/paralex_ontology.xml#lexeme
- `cell` (`string`): Reference to a cell identifier. The set of feature values as would appear in a gloss, separated by dots, eg. prs.ind.1sg or f.pl
    - constraints: a `cell` is obligatory.
    - `rdfProperty`: https://www.paralex-standard.org/paralex_ontology.xml#cell
- `phon_form` (`string`): Inflected form (phonemic or phonetic). The form, given in phonemic or phonetic notation, with sounds separated by spaces
    - `rdfProperty`: https://www.paralex-standard.org/paralex_ontology.xml#phon_form
    - `missingValues`: `#DEF#`

- **orth_form** (`string`): Inflected form (orthographic). The form, given orthographically

  - **rdfProperty**: https://www.paralex-standard.org/paralex_ontolog y.xml#orth_form
  - **missingValues**: `#DEF#`

- **analysed_phon_form** (`string`): Inflected form with analysis, such as segmentation markers (phonemic or phonetic). The form, given in phonemic or phonetic notation, with sounds separated by spaces, and analysis markers.

  - **rdfProperty**: https://www.paralex-standard.org/paralex_ontolog y.xml#analysed_phon_form
  - **missingValues**: `#DEF#`

- **analysed_orth_form** (`string`): Inflected form with analysis, such as segmentation markers (orthographic). The form, given orthographically, with markers for analysis.

  - **rdfProperty**: https://www.paralex-standard.org/paralex_ontolog y.xml#analysed_orth_form
  - **missingValues**: `#DEF#`

- **frequency** (`number`): Frequency. Frequency for this row.

  - **rdfProperty**: https://www.paralex-standard.org/paralex_ontolog y.xml#frequency

- **analysis_tag** (`string`): Tags for marking separate analyses. Identifies sets of forms which are related by the same analysis. Eg: forms from two distinct sources, or a more phonetic and a more phonological transcription.

  - **rdfProperty**: https://www.paralex-standard.org/paralex_ontolog y.xml#analysis_tag

- **defectiveness_tag** (`string`): Tags for defectiveness status. Identifies sets of defective forms (eg. pluralia tantum).

  - **rdfProperty**: https://www.paralex-standard.org/paralex_ontolog y.xml#defectiveness_tag

- **epistemic_tag** (`string`): Tags for epistemic status. Identifies sets of forms with the same epistemic status.

  - **rdfProperty**: https://www.paralex-standard.org/paralex_ontolog y.xml#epistemic_tag

- **variants_tag** (`string`): Tags for form variants. Identifies sets of forms used by specific groups of speakers. Eg. dialectal variants.

  - **rdfProperty**: https://www.paralex-standard.org/paralex_ontolog y.xml#variants_tag

- **overabundance_tag** (`string`): Tags for overabundant forms. Identifies sets of overabundant forms. For example, overabundant forms across lexemes might belong to a series of regular and irregular forms, or a series of short and long forms, etc.

  - **rdfProperty**: https://www.paralex-standard.org/paralex_ontology.xml#overabundance_tag

- **comment** (`string`): Comment. Human-readable comment.

  - **rdfProperty**: http://www.w3.org/2000/01/rdf-schema#comment

- **source** (`string`): Source. Reference to a specific source (bibtex key). If used, the dataset should comprise a .bib file where the keys are referenced.

  - **rdfProperty**: https://www.paralex-standard.org/paralex_ontology.xml#source

- **POS** (`string`): Part of Speech. The relevant part of speech for this item. This must refer to a PartOfSpeech entity from the lexinfo (https://lexinfo.net/) ontology.

  - constraints: a `POS` must be one of the values: `verb`, `numeral`, `conjunction`, `noun`, `adposition`, `determiner`, `article`, `adverb`, `pronoun`, `fusedPreposition`, `adjective`, `symbol`, `particle`, `conditionalParticle`, `demonstrativePronoun`, `interjection`, `semiColon`, `diminutiveNoun`, `possessivePronoun`, `prepositionalAdverb`, `compoundPreposition`, `interrogativeRelativePronoun`, `possessiveParticle`, `plainVerb`, `letter`, `interrogativeDeterminer`, `relativePronoun`, `postposition`, `fusedPronounAuxiliary`, `interrogativeOrdinalNumeral`, `indefiniteOrdinalNumeral`, `strongPersonalPronoun`, `possessiveRelativePronoun`, `ordinalAdjective`, `collectivePronoun`, `commonNoun`, `infinitiveParticle`, `comparativeParticle`, `partitiveArticle`, `invertedComma`, `lightVerb`, `emphaticPronoun`, `distinctiveParticle`, `genericNumeral`, `possessiveAdjective`, `reflexivePossessivePronoun`, `colon`, `coordinationParticle`, `presentParticipleAdjective`, `fusedPrepositionPronoun`, `cardinalNumeral`, `indefiniteDeterminer`, `numeralFraction`, `questionMark`, `generalAdverb`, `superlativeParticle`, `point`, `indefiniteMultiplicativeNumeral`, `comma`, `closeParenthesis`, `futureParticle`, `personalPronoun`, `reflexivePersonalPronoun`, `adverbialPronoun`, `reciprocalPronoun`, `openParenthesis`, `pastParticipleAdjective`, `negativePronoun`, `relativeDeterminer`, `existentialPronoun`, `pronominalAdverb`, `relativeParticle`, `exclamativeDeterminer`, `multiplicativeNumeral`, `reflexiveDeterminer`, `modal`, `unclassifiedParticle`, `properNoun`, `allusivePronoun`, `interrogativeCardinalNumeral`, `bullet`, `subordinatingConjunction`, `irreflexivePersonalPronoun`, `possessiveDeterminer`, `negativeParticle`, `indefinitePronoun`, `generalizationWord`, `coordinatingConjunction`, `deficientVerb`, `adjective-i`, `impersonalPronoun`, `indefiniteCardinalNumeral`,

```
adjective-na,    qualifierAdjective,    affirmativeParticle,
mainVerb,    fusedPrepositionDeterminer,    indefiniteArticle,
weakPersonalPronoun,suspensionPoints,interrogativeMultiplicativeNumeral,
affixedPersonalPronoun, auxiliary, circumposition, copula,
demonstrativeDeterminer,participleAdjective,exclamativePoint,
interrogativePronoun,    presentativePronoun,    punctuation,
definiteArticle,    slash,    exclamativePronoun,    preposition,
conditionalPronoun, relationNoun, interrogativeParticle.
```

- **rdfProperty**: https://www.paralex-standard.org/paralex_ontolog y.xml#POS

frequencies

Frequencies - This file is located in `frequencies.csv`.

- The identifier column (or `primaryKey`) is `['freq_id']`

- **Formal relations (foreignKeys)** with other tables:

Each value in column	Must refer to
`['cell']`	`['cell_id']` in table `cells`
`['form']`	`['form_id']` in table `forms`
`['lexeme']`	`['lexeme_id']` in table `lexemes`

Columns defined by `frequencies-schema`:

- **`freq_id`** (`string`): Frequency record identifier. One frequency value, for a single data point and source

    - constraints: a `freq_id` is obligatory; it must be unique.

- **`form`** (`string`): Reference to form table row identifiers. These identifiers are specific to form, lexeme, cell triples.

    - **rdfProperty**: https://www.paralex-standard.org/paralex_ontolog y.xml#form

- **`lexeme`** (`string`): Reference to a lexeme identifier. Lexeme identifiers must be unique to paradigms.

    - **rdfProperty**: https://www.paralex-standard.org/paralex_ontolog y.xml#lexeme

- **`cell`** (`string`): Reference to a cell identifier. The set of feature values as would appear in a gloss, separated by dots, eg. prs.ind.1sg or f.pl

    - **rdfProperty**: https://www.paralex-standard.org/paralex_ontolog y.xml#cell

- **frequency** (`number`): Frequency. Frequency for this row.

  - **rdfProperty**: https://www.paralex-standard.org/paralex_ontology.xml#frequency

- **analysis_tag** (`string`): Tags for marking separate analyses. Identifies sets of forms which are related by the same analysis. Eg: forms from two distinct sources, or a more phonetic and a more phonological transcription.

  - **rdfProperty**: https://www.paralex-standard.org/paralex_ontology.xml#analysis_tag

- **defectiveness_tag** (`string`): Tags for defectiveness status. Identifies sets of defective forms (eg. pluralia tantum).

  - **rdfProperty**: https://www.paralex-standard.org/paralex_ontology.xml#defectiveness_tag

- **epistemic_tag** (`string`): Tags for epistemic status. Identifies sets of forms with the same epistemic status.

  - **rdfProperty**: https://www.paralex-standard.org/paralex_ontology.xml#epistemic_tag

- **variants_tag** (`string`): Tags for form variants. Identifies sets of forms used by specific groups of speakers. Eg. dialectal variants.

  - **rdfProperty**: https://www.paralex-standard.org/paralex_ontology.xml#variants_tag

- **overabundance_tag** (`string`): Tags for overabundant forms. Identifies sets of overabundant forms. For example, overabundant forms across lexemes might belong to a series of regular and irregular forms, or a series of short and long forms, etc.

  - **rdfProperty**: https://www.paralex-standard.org/paralex_ontology.xml#overabundance_tag

- **source** (`string`): Source. Reference to a specific source (bibtex key). If used, the dataset should comprise a .bib file where the keys are referenced.

  - **rdfProperty**: https://www.paralex-standard.org/paralex_ontology.xml#source

sounds

Sound inventory with distinctive features - This file is located in `sounds.csv`.

- The identifier column (or `primaryKey`) is `['sound_id']`
- `missingValues`: "

Columns defined by `sounds-schema`:

- **sound_id** (`string`): sound representation. These identifiers are specific to sounds.

    - constraints: a `sound_id` is obligatory; it must be unique.

- **label** (`string`): label for this row. A human readable label for the row.

    - **rdfProperty**: http://www.w3.org/2000/01/rdf-schema#label

- **comment** (`string`): Comment. Human-readable comment.

    - **rdfProperty**: http://www.w3.org/2000/01/rdf-schema#comment

- **CLTS_id** (`string`): Identifier of this sound in CLTS. Reference to this sound in CLTS data.

    - constraints: a `CLTS_id` must be unique.

    - **rdfProperty**: https://www.paralex-standard.org/paralex_ontolog y.xml#CLTS_id

- **PHOIBLE_id** (`string`): Identifier of this sound in PHOIBLE. Reference to this sound in PHOIBLE.

    - constraints: a `PHOIBLE_id` must be unique.

    - **rdfProperty**: https://www.paralex-standard.org/paralex_ontolog y.xml#PHOIBLE_id

graphemes

Graphemes inventory - This file is located in `graphemes.csv`.

- The identifier column (or `primaryKey`) is `['grapheme_id']`
- `missingValues`: "

Columns defined by `graphemes-schema`:

- **grapheme_id** (`string`): grapheme representation. These identifiers are specific to graphemes.

    - constraints: a `grapheme_id` is obligatory; it must be unique.

    - **rdfType**: http://linguistics-ontology.org/gold/2010/Grapheme

- **comment** (`string`): Comment. Human-readable comment.

    - **rdfProperty**: http://www.w3.org/2000/01/rdf-schema#comment

- **canonical_order** (`integer`): Sorting order for visual presentation. The order in which items are canonically presented. Use integers to represent relative order, order is used per-item.

    - **rdfProperty**: https://www.paralex-standard.org/paralex_ontolog y.xml#canonical_order

cells

Paradigm cells - This file is located in `cells.csv`.

- The identifier column (or `primaryKey`) is `['cell_id']`

Columns defined by `cells-schema`:

- `cell_id` (`string`): Cell identifier. The set of feature values as would appear in a gloss, separated by dots, eg. prs.ind.1sg or f.pl

  – constraints: a `cell_id` is obligatory; it must be unique.

- `label` (`string`): label for this row. A human readable label for the row.

  – **rdfProperty**: http://www.w3.org/2000/01/rdf-schema#label

- `unimorph` (`string`): Cell in unimorph format. The cell, written following the unimorph schema

  – **rdfProperty**: https://www.paralex-standard.org/paralex_ontology.xml#unimorph

- `ud` (`string`): Cell in the universal dependency format. The cell, written following the universal dependency format

  – **rdfProperty**: https://www.paralex-standard.org/paralex_ontology.xml#ud

- `comment` (`string`): Comment. Human-readable comment.

  – **rdfProperty**: http://www.w3.org/2000/01/rdf-schema#comment

- `POS` (`string`): Part of Speech. The relevant part of speech for this item. This must refer to a PartOfSpeech entity from the lexinfo (https://lexinfo.net/) ontology.

  – constraints: a `POS` must be one of the values: verb, numeral, conjunction, noun, adposition, determiner, article, adverb, pronoun, fusedPreposition, adjective, symbol, particle, conditionalParticle, demonstrativePronoun, interjection, semiColon, diminutiveNoun, possessivePronoun, prepositionalAdverb, compoundPreposition, interrogativeRelativePronoun, possessiveParticle, plainVerb, letter, interrogativeDeterminer, relativePronoun, postposition, fusedPronounAuxiliary, interrogativeOrdinalNumeral, indefiniteOrdinalNumeral, strongPersonalPronoun, possessiveRelativePronoun, ordinalAdjective, collectivePronoun, commonNoun, infinitiveParticle, comparativeParticle, partitiveArticle, invertedComma, lightVerb, emphaticPronoun, distinctiveParticle, genericNumeral, possessiveAdjective, reflexivePossessivePronoun, colon, coordinationParticle, presentParticipleAdjective, fusedPrepositionPronoun, cardinalNumeral, indefiniteDeterminer, numeralFraction, questionMark, generalAdverb, superlativeParticle, point,

indefiniteMultiplicativeNumeral, comma, closeParenthesis,
futureParticle, personalPronoun, reflexivePersonalPronoun,
adverbialPronoun, reciprocalPronoun, openParenthesis,
pastParticipleAdjective, negativePronoun, relativeDeterminer,
existentialPronoun, pronominalAdverb, relativeParticle,
exclamativeDeterminer, multiplicativeNumeral, reflexiveDeterminer,
modal, unclassifiedParticle, properNoun, allusivePronoun,
interrogativeCardinalNumeral, bullet, subordinatingConjunction,
irreflexivePersonalPronoun, possessiveDeterminer, negativeParticle,
indefinitePronoun, generalizationWord, coordinatingConjunction,
deficientVerb, adjective-i, impersonalPronoun, indefiniteCardinalNumeral,
adjective-na, qualifierAdjective, affirmativeParticle,
mainVerb, fusedPrepositionDeterminer, indefiniteArticle,
weakPersonalPronoun, suspensionPoints, interrogativeMultiplicativeNumeral,
affixedPersonalPronoun, auxiliary, circumposition, copula,
demonstrativeDeterminer, participleAdjective, exclamativePoint,
interrogativePronoun, presentativePronoun, punctuation,
definiteArticle, slash, exclamativePronoun, preposition,
conditionalPronoun, relationNoun, interrogativeParticle.

- – **rdfProperty**: https://www.paralex-standard.org/paralex_ontology.xml#POS

- **frequency** (`number`): Frequency. Frequency for this row.

  - – **rdfProperty**: https://www.paralex-standard.org/paralex_ontology.xml#frequency

- **canonical_order** (`integer`): Sorting order for visual presentation. The order in which items are canonically presented. Use integers to represent relative order, order is used per-item.

  - – **rdfProperty**: https://www.paralex-standard.org/paralex_ontology.xml#canonical_order

features-values

Grammatical features values - This file is located in `features-values.csv`.

- The identifier column (or `primaryKey`) is `['value_id']`

Columns defined by `features-values-schema`:

- **value_id** (`string`): Grammatical Feature value identifier. Identifier for the grammatical feature value (as found in the cell)

  - – constraints: a `value_id` is obligatory; it must be unique.

- **label** (`string`): label for this row. A human readable label for the row.

  - – **rdfProperty**: http://www.w3.org/2000/01/rdf-schema#label

- **feature** (`string`): feature. The name of the dimension of this feature, eg. case, tense, modality, voice, force, gender, evidentiality, person, number, polarity…

  – constraints: a `feature` is obligatory.

  – `rdfProperty`: https://www.paralex-standard.org/paralex_ontology.xml#feature

- **comment** (`string`): Comment. Human-readable comment.

  – `rdfProperty`: http://www.w3.org/2000/01/rdf-schema#comment

- **POS** (`string`): Part of Speech. The relevant part of speech for this item. This must refer to a PartOfSpeech entity from the lexinfo (https://lexinfo.net/) ontology.

  – constraints: a `POS` must be one of the values: `verb`, `numeral`, `conjunction`, `noun`, `adposition`, `determiner`, `article`, `adverb`, `pronoun`, `fusedPreposition`, `adjective`, `symbol`, `particle`, `conditionalParticle`, `demonstrativePronoun`, `interjection`, `semiColon`, `diminutiveNoun`, `possessivePronoun`, `prepositionalAdverb`, `compoundPreposition`, `interrogativeRelativePronoun`, `possessiveParticle`, `plainVerb`, `letter`, `interrogativeDeterminer`, `relativePronoun`, `postposition`, `fusedPronounAuxiliary`, `interrogativeOrdinalNumeral`, `indefiniteOrdinalNumeral`, `strongPersonalPronoun`, `possessiveRelativePronoun`, `ordinalAdjective`, `collectivePronoun`, `commonNoun`, `infinitiveParticle`, `comparativeParticle`, `partitiveArticle`, `invertedComma`, `lightVerb`, `emphaticPronoun`, `distinctiveParticle`, `genericNumeral`, `possessiveAdjective`, `reflexivePossessivePronoun`, `colon`, `coordinationParticle`, `presentParticipleAdjective`, `fusedPrepositionPronoun`, `cardinalNumeral`, `indefiniteDeterminer`, `numeralFraction`, `questionMark`, `generalAdverb`, `superlativeParticle`, `point`, `indefiniteMultiplicativeNumeral`, `comma`, `closeParenthesis`, `futureParticle`, `personalPronoun`, `reflexivePersonalPronoun`, `adverbialPronoun`, `reciprocalPronoun`, `openParenthesis`, `pastParticipleAdjective`, `negativePronoun`, `relativeDeterminer`, `existentialPronoun`, `pronominalAdverb`, `relativeParticle`, `exclamativeDeterminer`, `multiplicativeNumeral`, `reflexiveDeterminer`, `modal`, `unclassifiedParticle`, `properNoun`, `allusivePronoun`, `interrogativeCardinalNumeral`, `bullet`, `subordinatingConjunction`, `irreflexivePersonalPronoun`, `possessiveDeterminer`, `negativeParticle`, `indefinitePronoun`, `generalizationWord`, `coordinatingConjunction`, `deficientVerb`, `adjective-i`, `impersonalPronoun`, `indefiniteCardinalNumeral`, `adjective-na`, `qualifierAdjective`, `affirmativeParticle`, `mainVerb`, `fusedPrepositionDeterminer`, `indefiniteArticle`, `weakPersonalPronoun`, `suspensionPoints`, `interrogativeMultiplicativeNumeral`, `affixedPersonalPronoun`, `auxiliary`, `circumposition`, `copula`, `demonstrativeDeterminer`, `participleAdjective`, `exclamativePoint`,

interrogativePronoun, presentativePronoun, punctuation, definiteArticle, slash, exclamativePronoun, preposition, conditionalPronoun, relationNoun, interrogativeParticle.

- **rdfProperty**: https://www.paralex-standard.org/paralex_ontology.xml#POS

- **unimorph** (`string`): Cell in unimorph format. The cell, written following the unimorph schema

  - **rdfProperty**: https://www.paralex-standard.org/paralex_ontology.xml#unimorph

- **ud** (`string`): Cell in the universal dependency format. The cell, written following the universal dependency format

  - **rdfProperty**: https://www.paralex-standard.org/paralex_ontology.xml#ud

- **canonical_order** (`integer`): Sorting order for visual presentation. The order in which items are canonically presented. Use integers to represent relative order, order is used per-item.

  - **rdfProperty**: https://www.paralex-standard.org/paralex_ontology.xml#canonical_order

lexemes

Lexemes - This file is located in `lexemes.csv`.

- The identifier column (or `primaryKey`) is `['lexeme_id']`

Columns defined by `lexemes-schema`:

- **lexeme_id** (`string`): Identifier for the lexeme. Lexeme identifiers. Often, they are identical to the label (lemma). However, they must be unique to paradigms, distinguishing homonyms with different inflection. For example, the animal mouse/mice and the computer peripheric mouse/mouses would both have the label 'mouse' but could be identified by the lexeme identifiers mouse_1 and mouse_2.

  - constraints: a `lexeme_id` is obligatory; it must be unique.

- **inflection_class** (`string`): Inflection class identifier. This identifier groups together lexemes of the same inflection class.

  - **rdfProperty**: https://www.paralex-standard.org/paralex_ontology.xml#inflection_class

- **source** (`string`): Source. Reference to a specific source (bibtex key). If used, the dataset should comprise a .bib file where the keys are referenced.

- – rdfProperty: https://www.paralex-standard.org/paralex_ontolog
  y.xml#source

- **frequency** (`number`): Frequency. Frequency for this row.

  - – rdfProperty: https://www.paralex-standard.org/paralex_ontolog
    y.xml#frequency

- **label** (`string`): label for this row. A human readable label for the row.

  - – rdfProperty: http://www.w3.org/2000/01/rdf-schema#label

- **meaning** (`string`): Definition for this lexeme. This is a description of the
  lexeme's overall meaning.

  - – rdfProperty: https://www.paralex-standard.org/paralex_ontolog
    y.xml#meaning

- **gloss** (`string`): Short meaning, used for glossing.. Gloss for this lexeme.

  - – rdfProperty: https://www.paralex-standard.org/paralex_ontolog
    y.xml#gloss

- **comment** (`string`): Comment. Human-readable comment.

  - – rdfProperty: http://www.w3.org/2000/01/rdf-schema#comment

- **POS** (`string`): Part of Speech. The relevant part of speech for this
  item. This must refer to a PartOfSpeech entity from the lexinfo
  (https://lexinfo.net/) ontology.

  - – constraints: a `POS` must be one of the values: `verb`, `numeral`,
    `conjunction`, `noun`, `adposition`, `determiner`, `article`, `adverb`,
    `pronoun`, `fusedPreposition`, `adjective`, `symbol`, `particle`,
    `conditionalParticle`, `demonstrativePronoun`, `interjection`,
    `semiColon`, `diminutiveNoun`, `possessivePronoun`, `prepositionalAdverb`,
    `compoundPreposition`, `interrogativeRelativePronoun`, `possessiveParticle`,
    `plainVerb`, `letter`, `interrogativeDeterminer`, `relativePronoun`,
    `postposition`, `fusedPronounAuxiliary`, `interrogativeOrdinalNumeral`,
    `indefiniteOrdinalNumeral`, `strongPersonalPronoun`, `possessiveRelativePronoun`,
    `ordinalAdjective`, `collectivePronoun`, `commonNoun`, `infinitiveParticle`,
    `comparativeParticle`, `partitiveArticle`, `invertedComma`,
    `lightVerb`, `emphaticPronoun`, `distinctiveParticle`, `genericNumeral`,
    `possessiveAdjective`, `reflexivePossessivePronoun`, `colon`,
    `coordinationParticle`, `presentParticipleAdjective`, `fusedPrepositionPronoun`,
    `cardinalNumeral`, `indefiniteDeterminer`, `numeralFraction`,
    `questionMark`, `generalAdverb`, `superlativeParticle`, `point`,
    `indefiniteMultiplicativeNumeral`, `comma`, `closeParenthesis`,
    `futureParticle`, `personalPronoun`, `reflexivePersonalPronoun`,
    `adverbialPronoun`, `reciprocalPronoun`, `openParenthesis`,
    `pastParticipleAdjective`, `negativePronoun`, `relativeDeterminer`,
    `existentialPronoun`, `pronominalAdverb`, `relativeParticle`,

exclamativeDeterminer, multiplicativeNumeral, reflexiveDeterminer, modal, unclassifiedParticle, properNoun, allusivePronoun, interrogativeCardinalNumeral, bullet, subordinatingConjunction, irreflexivePersonalPronoun, possessiveDeterminer, negativeParticle, indefinitePronoun, generalizationWord, coordinatingConjunction, deficientVerb, adjective-i, impersonalPronoun, indefiniteCardinalNumeral, adjective-na, qualifierAdjective, affirmativeParticle, mainVerb, fusedPrepositionDeterminer, indefiniteArticle, weakPersonalPronoun, suspensionPoints, interrogativeMultiplicativeNumeral, affixedPersonalPronoun, auxiliary, circumposition, copula, demonstrativeDeterminer, participleAdjective, exclamativePoint, interrogativePronoun, presentativePronoun, punctuation, definiteArticle, slash, exclamativePronoun, preposition, conditionalPronoun, relationNoun, interrogativeParticle.

- – `rdfProperty`: https://www.paralex-standard.org/paralex_ontology.xml#POS

- **`language_ID`** (`string`): Identifier for the language. Language identifiers should use some standard ID (iso code, glottocode, etc)

  - – `rdfProperty`: https://www.paralex-standard.org/paralex_ontology.xml#language_ID

- **`analysis_tag`** (`string`): Tags for marking separate analyses. Identifies sets of forms which are related by the same analysis. Eg: forms from two distinct sources, or a more phonetic and a more phonological transcription.

  - – `rdfProperty`: https://www.paralex-standard.org/paralex_ontology.xml#analysis_tag

- **`defectiveness_tag`** (`string`): Tags for defectiveness status. Identifies sets of defective forms (eg. pluralia tantum).

  - – `rdfProperty`: https://www.paralex-standard.org/paralex_ontology.xml#defectiveness_tag

- **`epistemic_tag`** (`string`): Tags for epistemic status. Identifies sets of forms with the same epistemic status.

  - – `rdfProperty`: https://www.paralex-standard.org/paralex_ontology.xml#epistemic_tag

- **`variants_tag`** (`string`): Tags for form variants. Identifies sets of forms used by specific groups of speakers. Eg. dialectal variants.

  - – `rdfProperty`: https://www.paralex-standard.org/paralex_ontology.xml#variants_tag

- **`overabundance_tag`** (`string`): Tags for overabundant forms. Identifies sets of overabundant forms. For example, overabundant forms across

lexemes might belong to a series of regular and irregular forms, or a series of short and long forms, etc.

- **rdfProperty**: https://www.paralex-standard.org/paralex_ontology.xml#overabundance_tag

tags

Tags mark rows which have commonalities - This file is located in `tags.csv`.

- The identifier column (or `primaryKey`) is `['tag_id']`

Columns defined by `tags-schema`:

- **tag_id** (`string`): Tag id. The label for a set of forms which have something in common.
    - constraints: a `tag_id` is obligatory; it must be unique.
- **tag_column_name** (`string`): Name of the tag column in the forms table. The name of the column this tag is used in the forms table
    - constraints: a `tag_column_name` is obligatory; it must match the regular expression `[^ ]+_tag`.
- **comment** (`string`): Comment. Human-readable comment.
    - **rdfProperty**: http://www.w3.org/2000/01/rdf-schema#comment

sources

Resources| Bibliographical references.

- This file is located in `sources.bib`.

readme

Read me| Basic documentation

- This file is located in `readme.md`.

data_sheet

Data Sheet| Data Sheet

- This file is located in `data_sheet.md`.