Error tags used in the Hebrew essay corpus

* This supplementary accompanies the Hebrew essay corpus. It refers to the error annotation scheme developed in Section 4.2 of:

Gafni, C., Herzig Sheinfux, L., Klunover, H., Prior, A., and Wintner, S. Analyzing learner language: the case of the Hebrew essay corpus. *Language Resources and Evaluation*.

* Corresponding author: Chen Gafni. E-mail: chen.gafni@gmail.com.

* Corpus website: https://github.com/HaifaCLG/HebrewEssayCorpus

In the proposed error annotation scheme, error tags have the general form of `function(arguments)`. *Functions* indicate the nature of the deviation between the original and revised token. *Arguments* to the functions usually denote linguistic categories affected by the error. This structure is described in Section 4.2.1 in Gafni et al. The lists of functions and linguistic categories are provided in the following sub-sections.[1]

Each function has a required number of arguments: 0, 1, or 2. Most functions require one argument. There are also special meta-analytic codes (##, &&, $$) that can be passed as arguments to any function. These codes represent cases of dependent corrections and annotated but uncorrected errors, which are discussed in Sections 4.2.3 and 4.2.4 in Gafni et al.

Error tags in the format mentioned above appear in columns titled “Error” in the annotated files. In addition, columns titled “Interp” contain additional higher-level, interpretive (or, explanatory) error tags that elaborate on the errors in the “Error” columns. Interpretive error tags are discussed in Section 4.2.5 in Gafni et al.

[1] In designing the annotation scheme, we consulted the categories and features defined within the framework of Universal Dependencies (Nivre et al., 2020) and specifically for Hebrew (see: https://universaldependencies.org/).

1. General functions

Three functions are general and can take most valid linguistic categories as arguments:

Function	Description	Example	
`miss`	Missing element	`miss(prep)`	(“missing preposition”)
`redun`	Redundant element	`redun(conj)`	(“redundant conjunction”)
`wrong`	Incorrect element	`wrong(plural)`	(“wrong plural form”)

2. Orthography

This sub-section describes tagging of various orthographic errors (e.g., spelling errors). Presumably, these errors reflect phonological and visual processing of the L2 orthography (i.e., Hebrew), as well as discrepancies between the L1 and L2 phono-orthographic systems. We used these tags when the likely cause of an error was orthographic or phonological.

2.1. Basic tagging

The basic orthographic category is a letter or sequence of letters, represented by `x` and `y` in the following descriptions. Annotation of orthographic errors typically uses functions with different names than the general functions described in Section 1 above. This variation was done in order to facilitate orthographic queries in the corpus (e.g., identifying all cases of missing letters as opposed to other linguistic categories).

`oMiss(x)`: Missing letter (or sequence of letters).

- Example: `oMiss(ו)` (=missing letter ו):
  Error: *טכנלוגיה*, correct: *טכנולוגיה* [texn**o**logya] ‘technology’.

`oRedun(x)`: Redundant letter (or sequence of letters).

- Example: `oRedun(ל)` (=redundant letter ל):
  Error: *ללהישאר*, correct: *להישאר* [lehiša’er] ‘stay.INF’.

`shouldB(x,y)`: Letter substitution: x should be y.

- Example: `shouldB(ב,מ)` (=the letter ב should be replaced by the letter מ):
  Error: *שובְרים* [šo**v**rim] ‘break.M.PL.PRS’, correct: *שומְרים* [šo**m**rim] ‘keep.M.PL.PRS’

`metathesis(x)`: Letter transposition: x is the target string.

- Example: `metathesis(על)` (=the sequence of letters על in the target hypothesis has been transposed):
  Error: *תולעת* [tola'at] 'worm', correct: *תועלת* [to'elet] 'benefit'.

Notes

1. A sequence of missing or redundant letters is usually tagged as separate errors, unless the sequence of letters represents a diphthong, which is considered a single sound. For example, the correction of the non-existent word *יוהם* to *הם* [hem] 'they' involves deletion of two letters, which is tagged as two separate errors: `oRedun(י)` and `oRedun(ו)`. By contrast, the correction of the non-existent word *להתפתייח* [hypothesized pronunciation: lehitpate**y**ax] to *להתפתח* [lehitpateax] 'evolve.INF' is tagged as a single error of a redundant diphthong: `oRedun(יי)`.

2. The `metathesis` code usually indicates transposition of two adjacent letters. Transposition of non-adjacent letters is tagged as if they were adjacent (e.g., error: *קבורות* [k**v**u**r**ot] 'burried.F.PL', correct: *קרובות* [k**r**o**v**ot] 'close.F.PL', tag: `metathesis(בר)`). When a sequence longer than two letters involves multiple letter transpositions, the entire target sequence is included in the code (e.g., error: *רחבה* [**r**e**x**a**v**a] 'wide.F.SG'/ [**r**a**x**a**v**a] 'yard', correct: *חברה* [**x**e**v**r**a**] 'society', tag: `metathesis(חברה)`).

2.2. Non-alphanumeric categories

Hebrew uses two non-alphanumeric symbols as orthographic entities. Geresh, an apostrophe-like sign, is used as a phonetic diacritic attached to letters, in order to represent sounds that Modern Hebrew borrowed from other languages:

Letter	**Sound**	**Geresh-modified letter**	**Modified sound**
ג	[g] (as in English g in ***g**o*)	ג׳	[j] (as in English j in ***j**ar*)
ז	[z] (as in English z in ***z**oo*)	ז׳	[ž] (as in English s in *mea**s**ure*)
צ	[c] (as in English ts in *ca**ts***)	צ׳	[č] (as in English ch in ***ch**eck*)

The Geresh is also used as a marker of abbreviation as in *פרופ׳* (short form of *פרופסור* 'professor'). A close relative of the Geresh is Gershayim (lit. double Geresh), which is used to mark abbreviations and acronyms, as in *ד״ר* (short form of *דוקטור* 'doctor') and *ח״כ* (acronym of *חבר כנסת* 'a member of the Israeli parliament').[2]

Similarly to the cases discussed above, orthographic errors may involve the use, non-use, or misuse of Geresh and Gershayim. We defined two categories for these elements, distinguishing the phonetic use of the Geresh from the use of these symbols as abbreviation markers. Note that since these symbols are not letters, they are normally tagged using the general functions (i.e., `miss`, `redun`, `wrong`) rather than orthographic specific functions (i.e., `oMiss`, `oRedun`, `shouldB`).

`abbrev`: Errors involving abbreviation markers.

- Example: `miss(abbrev)` (=missing abbreviation):
  Error: *חול* [xol] 'sand', correct: *חו״ל* [xul] 'abroad' (abbreviation of *חוץ לארץ*, lit.: 'out to the country')

`phonApos`: Errors involving the phonetic Geresh.

- Example: `redun(phonApos)` (=redundant phonetic Geresh):
  Error: *וואצ׳אפ* [wačap], correct: *וואצאפ* [wacap] 'WhatsApp'

Note that, when a phonetic Geresh is inserted or deleted and the hosting letter is also changed, the error will be tagged using the `shouldB` since such errors are regarded as a sequence error rather than separate letter and Geresh errors. For example, the spelling correction *סנאפשט* [snepšet] → *סנאפצ׳ט* [snepčet] 'Snapchat' is tagged `shouldB(ש,'צ)` (such an error was found in several essays authored by L1 French speakers, and probably reflects the fact that the language lacks the phoneme /č/).

[2] Geresh and Gershayim are also used as single and double quotation markers, respectively.

2.3. Holistic errors

Occasionally, an orthographic error could not be analyzed at the individual-letter level, but rather seemed to affect the entire word. Such cases probably reflect a deep level of confusion or transfer from another language. Annotation of these errors use the general function `wrong` with single argument denoting the affected property.

`wrong(HebForm)`: Non-standard borrowing ("Hebraization") of a non-Hebrew word or morphology.

- Example: Error: *סולוציאות* (hypothesized pronunciation: [soluciot]): probably a non-standard borrowed form of *solutions*, where an equivalent standard Hebrew form exists. Correct: *פתרונות* [pitronot].

`wrong(script)`: Non-Hebrew script.

- Example: Error: *Face to face*: an equivalent Hebrew phrase could be used.
  Correct: *פנים אל פנים* [panim el panim].

`wrong(spell)`: Accumulation of spelling errors that disrupts the entire word structure.

- Example: *אהבירת*: likely a deformation of *ארה"ב* 'USA'.

2.4. Higher-level interpretations

Orthographic errors are always tagged according to the surface difference between the author version and the target hypothesis. Occasionally, the error can be explained in terms of phonological or visual processing, with some confidence. Such cases can be tagged using an appropriate explanatory code from the list below. Since the explanatory codes represent a rather speculative analysis, they are never used alone. Rather, they are indicated in separate columns labelled "Interp".

`homophone`: A homophonic spelling error: substitution of a letter representing the same sound. This tag is used as an interpretation of the `shouldB` function.

- Example: main error tag: `shouldB(ק,כ)` (=the letter ק should be replaced by כ, both representing the sound [k]): Error: *קסף*, Correct: *כסף* [kesef] 'money'.

`pronuncL2`: A spelling error that likely stems from the typical pronunciation of L2 speakers with a specific L1 (using this code requires knowledge of the phonological system of the relevant L1).

- Example: main error tag: `shouldB(ב,פ)` (=the letter ב should be replaced by פ, the error likely represents the typical pronunciation of L2 speakers): Error: *לשבר* [leša**b**er], Correct: *לשפר* [leša**p**er] 'improve.INF'. This specific letter substitution is typical of L1 Arabic speakers, which has /b/ as a phoneme, but not /p/.

`pronuncReg`: A spelling error that reflects the typical pronunciation of the word (by either L1 or L2 speakers).

- Example: main error tag: `oRedun(י)` (=a redundant letter י, which represents the vowel [i]): Error: *עים*, Correct: *עם* [im] 'with'.

`finPos`: A letter should appear in its final form in a word-final position. This tag is used as an interpretation of the `shouldB` function.

- Example: main error tag: `shouldB(פ,ף)` (=the letter פ should be replaced by its final form version when in a word-final position):
  Error: *שאפ* (hypothesized pronunciation: [ša'a**p**]), Correct: *שאף* [ša'a**f**] 'aspire.3SG.M.PST'.

`nonFinPos`: A letter should appear in its non-final form when not in a word-final position. This tag is used as an interpretation of the `shouldB` function.

- Example: main error tag: `shouldB(ף,פ)` (the letter ף should be replaced by its non-final form version when not in a word-final position):
  Error: *מעדיף**ים***, Correct: *מעדי**פים*** [ma'adi**f**im] 'prefer.PL.M.PRS'.

`visual`: A substitution of similarly looking letters (especially in cursive Hebrew). This tag is used as an interpretation of the `shouldB` function.

- Example: main error tag: `shouldB(ף,ץ)` (the letter ף should be replaced by ץ, the two letters have similar visual appearance. The similarity is in cursive Hebrew):
  Error: *ארף* (cursive:  , hypothesized pronunciation: [aref]),
  Correct: *ארץ* (cursive: ) [arec] 'country'.

`analogy`: Orthographic error due to phonological and semantic similarity of a sub-word string to an existing word.

- Example: main error tag: `oRedun(א)` (=the letter א is redundant and was used possibly due to analogy to an existing word):
  Error: *מישהוא*, Correct: *מישהו* [mišehu] 'someone.M'. Possible cause for error: analogy to the word *הוא* [hu] 'he'.

`similar`: and error due to orthographic similarity between consonantal roots of author-version and target hypothesis (semantic similarity or preservation of linear sequence of letters are not necessary).

- Example: main error tag: `shouldB(א,נ)` (=the letter א should be replaced by נ. Possible reason for error: similarity to consonantal root in the target hypothesis):
  Error: *יביאו* [yavi'u] 'bring.3PL.M.FUT' (root: *ב.ו.א*),
  Correct: *יבינו* [yavi**n**u] 'understand.3PL.M.FUT' (root: *ב.י.נ*).

3. Morphology

3.1. Tokenization

Two error functions are used for tagging tokenization errors. Both functions are usually assigned to two consecutive tokens. One instance of the function does not take an argument, while the other takes the dummy token && as argument in places where the dummy token is inserted during annotation.

`split()`: A word is incorrectly split into two tokens.

- Example: Error: *ה צעירים* [ha ce'irim'], Correct: *הצעירים* [ha-ce'irim] 'the young' (in Hebrew, the definite article is prefixed to the modified word). Tagging:

Token	**Token_revised**	**Error1**
ה	&&	`split(&&)`
צעירים	הצעירים	`split()`

`merge()`: Two tokens are incorrectly merged into a single token.

- Error: *למישיש*, Correct: *למי שיש* [lemi še-yeš] 'to anyone who has'. Tagging:

Token	**Token_revised**	**Error1**
למישיש	למי	`merge()`
&&	שיש	`merge(&&)`

3.2. Non-linear patterns

This section describes a single code, which refers to an inappropriate combinations of consonantal roots and non-linear morphological patterns. The root consonants are represented by boldfaced letters in the Hebrew script and by uppercase letters in the phonetic transcription.

`wrong(pattern)`: An incorrect morphological pattern combined with a relevant consonantal root. The result could be an existing but irrelevant word, or a non-existing word.

- Example: Error: *להספיק* [lehaSPiK] 'suffice.INF', Correct: *לספק* [leSaPek] 'provide.INF'.
- Example: Error: *אשירים* [aŠiRim], Correct: *מאושרים* [me'uŠaRim] 'happy.PL.M'.

3.3. Linear morphology

The categories in this section refer to errors in linear stem-affix sequences. There are three main categories of linear morphology, which can be passed as arguments to all three general functions (i.e., `miss`, `redun`, `wrong`).

`plural`: Errors related to plural suffixes.

- Example: `redun(plural)` (=redundant plural suffix):
  Error: *אנחנו יכולים לדבר איתם **בטלפונים*** [anaxnu yexolim ledaber itam ba-**telefonim**] 'we can talk to them on the **telephones**', Correct: *אנחנו יכולים לדבר איתם **בטלפון*** [anaxnu yexolim ledaber itam ba-**telefon**] 'we can talk to them on the **telephone**'.

`possess`: Errors related to possessive suffixes.

- Example: `redun(possess)` (=redundant possessive suffix):
  Error: *יש חסרונותי**ה*** [yeš xesronot**ea**] 'has disadvantages **of her**',
  Correct: *יש חסרונות* [yeš xesronot] 'has disadvantages'.

`deriv`: Errors related to general derivative suffixes.

- Example: `wrong(deriv)` (=wrong derivative suffix):
  Error: *ספורט**יבים*** [sport**iv**im] 'sporty.PL.M', Correct: *ספורט**אים*** [sport**a'**im] 'athlete.PL.M'.

In addition, two specific codes refer to errors in inflected prepositions (in Hebrew, prepositions are combined with a following pronoun and, some prepositions are also combined with a following definite article).

`wrong(prep,infl)`: Incorrect inflection of a preposition due to inconsistencies in the inflection paradigm.

- Example: Error: *ממנם* [hypothesized pronunciation: mimenam],
  Correct: *מהם* [mehem] 'from them' (compare: *ממני* [mimeni] 'from me').

`mSplit(defPP)`: Incorrect in-word split of a definite preposition (inappropriate use of explicit definite article).

- Example: Error: *ב**ה**תקופה* [**be-ha**-tkufa], Correct: *ב**ת**קופה* [**ba**-tkufa] 'in the period (of time)'.

4. Syntax and lexicon

4.1. Lexical categories

This section describes various lexical categories, which cover both content words and function words. Each category can be combined with all three general functions (i.e., `miss`, `redun`, `wrong`), with only a few exceptions.

`conj`: Conjunctions: coordinating (e.g., ו [ve] 'and') and subordinating conjunctions (e.g., ש [še] 'that').

- Example: `miss(conj)` (=missing conjunction):
  Error: *לאנשים זקוקים לכך* [le-anašim zkukim lexax] 'to people need that',
  Correct: *לאנשים **ש**זקוקים לכך* [le-anašim **še**-zkukim lexax] 'to people **that** need that'.

`constr`: Construct state: juxtaposition of two nouns denoting a connection between them (e.g., possession), could be separated by the preposition *של* [šel] 'of'.

- Example: `redun(constr)` (=redundant inflection of the head noun in a non-construct state phrase): Error: *באוניברסיטת בחו"ל* [be-oniversita**t** bexul] 'in a university.**CONSTR** abroad', Correct: *באוניברסיטה בחו"ל* [be-oniversita bexul] 'in a university abroad'.

`cop`: Copulas: elements (mainly pronouns) connecting a nominal predicate with its subject.

- Example: `miss(cop)` (=a missing copula):
  Error: *הכסף לא המטרה* [ha-kesef lo ha-matara] 'the money not the goal',
  Correct: *הכסף **הוא** לא המטרה* [ha-kesef **hu** lo ha-matara] 'the money **is** not the goal'.

`dem`: Demonstratives (can be considered a sub-type of pronouns).

- Example: `wrong(dem)` (=wrong demonstrative): error: *בגלל **כך** הם שוכחים* [biglal **kax** hem šoxaxim] 'because of **that way**, they forget', correct: *בגלל **זה** הם שוכחים* [biglal **ze** hem šoxaxim] 'because of **that**, they forget'.

`det`: Determiners: the definite article ה /ha/, quantifiers (e.g., *כל* [kol] 'every') and interrogative determiners that precede the modified word (e.g., *איזה* [eyze] 'which').

- Example: `miss(det)` (=missing determiner):

Error: *דבר הכי חשוב* [davar haxi xašuv] ‘most important thing’,

Correct: *הדבר הכי חשוב* [**ha-**davar haxi xašuv] ‘**the** most important thing’.

`existential`: Existential predicates (mainly the existential quantifiers: *יש* [yeš] ‘there is/are’ and *אין* [eyn] ‘there isn’t/aren’t’).

- Example: `miss(existential)` (=a missing existential predicate):

  Error: *לכל אדם חלום* [le-kol adam xalom] ‘every man a dream’,

  Correct: *לכל אדם **יש** חלום* [le-kol adam **yeš** xalom] ‘every man **has** a dream’.

`lex`: Lexical item (content words): nouns, verb, adjectives. This category also covers selection of an incorrect consonant root in words conjugated in a non-linear pattern (errors in the pattern itself are analyzed using the `pattern` category; see Section 3.2 above).

- Example: `wrong(lex)` (=wrong lexical item):

  Error: *מצד **ראשון*** [micad **rišon**] ‘on the **first** hand’,

  Correct: *מצד **אחד*** [micad **exad**] ‘on the **one** hand’.

`neg`: Negation.

- Example: `wrong(neg)` (=incorrect negation element):

  Error: *זה **אי** אפשרי* [ze **i** efšari] ‘this is **un**possible’,

  Correct: *זה **בלתי** אפשרי* [ze **bilti** efšari] ‘this is **im**possible’.

`prep`: Prepositions.

- Example: `redun(prep)` (=redundant preposition):

  Error: *מכירה **עם** אנשים* [mekira **im** anašim] ‘know.F.SG.PRS **with** people’,

  Correct: *מכירה אנשים* [mekira anašim] ‘know.F.SG.PRES people’.

`pronoun`: Pronouns.

- Example: `wrong(pronoun)` (=wrong pronoun):

  Error: *הטכנולוגיה עוזרת לנו ליצור קשר עם **הכל*** [ha-texnologya ozeret lanu licor kešer im **hakol**] ‘technology helps communicate with **everything**’,

  Correct: *הטכנולוגיה עוזרת לנו ליצור קשר עם **כולם*** [ha-texnologya ozeret lanu licor kešer im **kulam**] ‘technology helps communicate with **everyone**’.

`punct`: Punctuation is not corrected in the current scheme, since it is not used consistently, even by native speakers. However, clear punctuation errors can be tagged. The error function should include $$ as argument to indicate that the error was not corrected.

- Example: `redun(punct,$$)` (=redundant punctuation):
  Error: *הטלוויזיה- משפיעה על אלפי אנשים* [ha-televizya**,** mašpi'a al alfey anašim] 'the TV**,** affects thousands of people',
  Correct (if punctuation had been corrected): *הטלוויזיה משפיעה על אלפי אנשים* [ha-televizya mašpi'a al alfey anašim] 'the TV affects thousands of people'.

`que`: Question words (not only in question sentences).

- Example: `redun(que)` (=redundant question word):
  Error: *דרכים יותר נוחות **איך** להתחבר* [draxim yoter noxot **ex** lehitxaber] 'more convenient ways **how** to connect',
  Correct: *דרכים יותר נוחות להתחבר* [draxim yoter noxot lehitxaber] 'more convenient ways to connect'.

`rel`: A component in a relative clause referring back to the antecedent in the main clause.

- Example: `miss(rel)` (=a missing referential component):
  Error: *הטכנולוגיה הפכה את העולם לכפר קטן **שכולם** יכולים לדבר עם כולם* [ha-texnologya hafxa et ha-olam lekfar katan še-kulam yexolim ledaber im kulam] 'technologoy turned the world into a small village which everyone can talk to everyone',
  Correct: *הטכנולוגיה הפכה את העולם לכפר קטן **שבו** כולם יכולים לדבר עם כולם* [ha-texnologya hafxa et ha-olam lekfar katan še-**bo** kulam yexolim ledaber im kulam] 'technologoy turned the world into a small village **in** which everyone can talk to everyone'.

4.2. Agreement

This section describes various agreement errors (in gender, person, number, or definiteness). The basic agreement tagging notation is `agree(cat1, cat2)`, where `cat1` and `cat2` denote the linguistic categories of the words that fail to agree in some feature.

`agree(det,noun)`: Agreement between a noun and its determiner (including quantifiers).

- Example: Error: *הרבה אפליקצי**ה*** [harbe aplikacy**a**] 'many application',
  Correct: *הרבה אפליקצי**ות*** [harbe aplikacy**ot**] 'many application**s**'.

`agree(nomReg,nomRect)`: Agreement between the nomen regens ("governing noun") and the nomen rectum ("governed noun") in a construct state.

- Example: Error: *כוח של המחשבה* [**koxo** šel ha-maxšava] 'power.GEN.SG.**M** of the thought (F)',
  Correct: *כוח**ה** של המחשבה* [**koxa** šel ha-maxšava] 'power.GEN.SG.**F** of the thought (F)'.

`agree(noun,adj)`: Agreement between a noun and a modifying adjective.

- Example: Error: *תמונות טוב**ים*** [tmunot **tovim**] 'good.**M**.PL picture(F).PL',
  Correct: *תמונות טוב**ות*** [tmunot **tovot**] 'good.**F**.PL picture(F).PL'.

`agree(noun,dem)`: Agreement between a noun and a deictic demonstrative.

- Example: Error: *השאלה הז**ה*** [ha-še'ela **haze**] 'this.**M** question(F)',
  Correct: *השאלה הז**ו*** [ha-še'ela **hazo**] 'this.**F** question(F)'.

`agree(noun,num)`: Agreement between a noun and a modifying numeral.

- Example: Error: *הפעם הראש**ון*** [ha-pa'am ha-**rišon**] 'the first.M time (F)',
  Correct: *הפעם הראש**ונה*** [ha-pa'am ha-**rišona**] 'the first.F time (F)'.

`agree(subj,cop)`: Agreement between the subject and the copula in a nominal clause.

- Example: Error: *רוב החדשות הם רעות* [rov ha-xadašot **hem** ra'ot] 'most of the news (F) are.**M** bad.F.PL',

Correct: *רוב החדשות **הן** רעות* [rov ha-xadašot **hen** ra'ot] 'most of the news (F) are.**F** bad.F.PL'.

`agree(subj,pred)`: Agreement between the subject and the predicate of a clause.

- Example: Error: *האם המשפט הזה **נכונה**?* [haim ha-mišpat haze **nexona**] 'is this sentence (M) **true (F)**?',
  Correct: *האם המשפט הזה **נכון**?* [haim ha-mišpat haze **naxon**] 'is this sentence (M) **true (M)**?'.

`agree(subj,existential)`: Agreement between the subject and an existential predicate of a clause (unlike other predicates, which normally appear after the subject, existential predicates normally appear before the subject).

- Example: Error: *קיימים מדינות* [kayam**im** medinot] 'there are.**M** countries (F)',
  Correct: *קיימות מדינות* [kayam**ot** medinot] 'there are.**F** countries (F)'.

`agree(component,anaphor)`: Agreement between a noun phrase or clause and its anaphor.

- Example: Error: *יש לו בעיות והוא צריך לפתור **אותם*** [yeš lo be'ayot ve-hu carix liftor **otam**] 'he has problems (F) and he needs to solve them.**M**',
  Correct: *יש לו בעיות והוא צריך לפתור **אותן*** [yeš lo be'ayot ve-hu carix liftor **otan**] 'he has problems (F) and he needs to solve them.**F**'.

Agreement errors occurring in constructions that do not fit into one of the above relations can be tagged using one of the following general tags:

`agree(general)`: Agreement between two components.

- Example: Error: *זה עשה את החיים **קלות*** [ze asa et ha-xaim **kalot**] 'it made life (M) **easy.F**',
  Correct: *זה עשה את החיים **קלים*** [ze asa et ha-xaim **kalim**] 'it made life (M) **easy.M**'.

`redun(agr)`: Redundant inflection of some component as if to express agreement with another component in the sentence.

- Example: Error: *אנשים הנמצאים **רחוקים*** [anašim hanimca'im **rexokim**] 'people who are **far.PL**',
  Correct: *אנשים הנמצאים **רחוק*** [anašim hanimca'im **raxok**] 'people who are **far.SG**'.

4.3. Argument structure

Missing arguments of a predicate (both subject and complements, i.e., objects) are tagged using the following codes. Elements that cannot be licensed as arguments of a predicate are tagged using one of the lexical categories' codes.[3]

`miss(subj)`: Missing subject.

- Example: Error: *יכולים לעשות הרבה דברים* [yexolim la'asot harbe dvarim] 'can.M.PL do many things',
  Correct: *אנחנו יכולים לעשות הרבה דברים* [**anaxnu** yexolim la'asot harbe dvarim] '**we** can.M.PL do many things'.

`miss(complement)`: Missing complement.

- Example: Error: *אם לא נעשה, אז זה לא יקרה* [im lo na'ase, az ze lo yikre] 'If we don't do, it won't happen',
- Correct: *אם לא נעשה משהו, אז זה לא יקרה* [im lo na'ase **mašehu**, az ze lo yikre] 'If we don't do **something**, it won't happen'.

Note: the `subj` and `complement` categories are used only with the `miss` function. Seemingly redundant arguments were never tagged as such, since arguments are licensed by syntactic positions and, by definition, an element cannot be licensed as an argument of a predicate, if the predicate has no free argument position. Instead, cases of so-called redundant arguments were tagged according to their corresponding lexical category (e.g., `redun(pronoun)`, standing for a redundant pronoun).

[3] Note that we use the term "argument" in two senses. In the most prominent use, "argument" is a feature of the annotation scheme. It refers to any linguistic category passed as an argument to a function that describes an error affecting the category. In the more restricted use, "argument" is a theoretical concept in linguistic theory. It refers to the syntactic relation between a predicate and its subject and complements.

4.4. Miscellaneous

Miscellaneous syntactic errors.

`wrong(dem,infl)`: Incorrect inflection of a demonstrative.[4]

- Example: Error: *זאת מה שלומדים* [**zot** ma še-lomdim] '**this.F** is what one learns',
  Correct: *זה מה שלומדים* [**ze** ma še-lomdim] '**this.M** is what one learns'.

`order()`: Incorrect word order.

- Example: Error: *הרבה זמן יותר* [harbe **zman yoter**] 'much **time more**',
  Correct: *הרבה יותר זמן* [harbe **yoter zman**] 'much **more time**'.

4.5. Higher-level interpretations

This section describes various explanatory, "high-level" tags that supplement tags of lexical and syntactic errors, including errors reflecting pragmatic and discourse issues.[5] The interpretive tags are indicated in separate columns labelled "Interp".

`irrelevant`: Wrong lexical item. The meaning is irrelevant to the context. This tag is used as an interpretation of the `wrong(lex)`tag.

- Example: Error: *שוכחים את רעד החיים* [šoxaxim et **ra'ad** ha-xaim] 'forgetting the **tremble** of life',
  Correct: tagging without correction: the intended meaning is unclear.

`word?`: Wrong lexical item – a non-existing word, unclear meaning.

- Example: Error: *אני רוצה להשעיל להם הצלחה* [ani roce **lahaš'il** lahem haclaxa] 'I want to **???** them success',
  Correct (hypothesized): *אני רוצה לאחל להם הצלחה* [ani roce **le'axel** lahem haclaxa] 'I want to **wish** them success'.

[4] This tag refers to the use of existing but contextually-inappropriate words. By contrast, the tag `wrong(prep,infl)` mentioned above refers to inflection errors resulting in non-existing words.

[5] Generally, pragmatic analysis and discourse analysis require context, which is partially unavailable in the Hebrew Essay Corpus, since the sentences of each essay were scrambled. However, some of the relevant information can be extracted based on thematic considerations and the internal coherence of individual sentences.

`infelicit`: Semantically infelicitous use. This tag is similar to `irrelevant` but is used with less severe errors, when there is an apparent semantic connection between the error and the target hypothesis.

- Example: Error: *אנשים **עושים** עדיפות לקריירה* [anašim **osim** adifut lakariyera] 'people **do** priority to the career',
  Correct: *אנשים **נותנים** עדיפות לקריירה* [anašim **notnim** adifut lakariyera] 'people **give** priority to the career'. Main error tag: `wrong(lex)`.

`colloc`: Deformation of a conventional collocation (a multi-word expression). The primary error could be related to any category (preposition, content word, etc.). However, it has an additional effect of disrupting a collocation, which can be analyzed as a lexical error.

- Example: Error:[6] *בצד אחד* [**be**'cad exad] '**from** the one hand',
  Correct: *מצד אחד* [**mi**'cad exad] '**on** the one hand'. Main error tag: `wrong(prep)` (although this is a preposition error, it also disrupts a conventional collocation).

`clause`: Agreement error across the boundary between a main clause and a subordinate clause.

- Example: Error: *יש הרבה תוכניות בטלוויזיה **המשודרים*** [yeš harbe toxniyot **hamešudarim** ba-televizya] 'there are many programs(F) which are broadcasted.**M**.PL on TV',
  Correct: *יש הרבה תוכניות בטלוויזיה **המשודרות*** [yeš harbe toxniyot **hamešudarot** ba-televizya] 'there are many programs(F) which are broadcasted.**F**.PL on TV'. Main error tag: `agree(subj,pred)`.

`generic`: Errors in the construction of a generic sentence.

- Example: Error: *חשוב להיות עם המשפחה **שלו*** [xašuv lihiyot im ha-mišpaxa **šelo**] 'it's important to be with **his** family',
  Correct: *חשוב להיות עם המשפחה **שלך*** [xašuv lihiyot im ha-mišpaxa **šelxa**] 'it's important to be with **your** family'. Main error tag: `wrong(pronoun)`.

[6] Collocation errors tend to be language-specific and are hard to reproduce accurately in translation. The translation provided in the given example reflects the effect of the error rather than the literal meaning of the words.

`cohерPerson`: Unexpected changes of the grammatical properties (person, gender, or number) of an entity in the course of a sentence.

- Example: Error: *אנחנו יכולים להתקשר לרופא או לדעת איפה הילדים שלי* [anaxnu yexolim lehitkašer la-rofe o lada'at eyfo ha-yeladim **šeli**] 'we can call the doctor or know where **my** children are', Correct: *אנחנו יכולים להתקשר לרופא או לדעת איפה הילדים שלנו* [anaxnu yexolim lehitkašer la-rofe o lada'at eyfo ha-yeladim **šelanu**] 'we can call the doctor or know where **our** children are'. Main error tag: `wrong(pronoun)`.

`cohерTense`: Unexpected changes of tense in the course of a sentence.

- Example: Error: *הוא ילמד ומצליח* [hu yilmad ve-**macliax**] 'he will learn and **succeeds**', Correct: *הוא ילמד ויצליח* [hu yilmad ve-**yacliax**] 'he will learn and **succeed**'.
  Main error tag: `wrong(pattern)`.

`register`: Inappropriate register (too low or too high/archaic).

- Example: Error: *אני לא יעשה את זה* [ani lo **ya'ase** et ze] 'I will not do.3SG.M that',
  Correct: *אני לא אעשה את זה* [ani lo **a'ase** et ze] 'I will not do.1SG that' (an agreement error originating in a phonological process that is common in colloquial Hebrew, but unacceptable in written register).
  Main error tag: `shouldB(י,א)`. In this case, the error can be analyzed as an effect of pronunciation. Thus, the interpretive tag would be a combination of both analyses: `prouncReg/register`.

References

Nivre, J., de Marneffe, M. C., Ginter, F., Haji, J., Manning, C. D., Pyysalo, S., Schuster, S., & Zeman, F. T. D. (2020). Universal dependencies v2: An evergrowing multilingual treebank collection. *LREC 2020 - 12th International Conference on Language Resources and Evaluation, Conference Proceedings*, *May*, 4034–4043.