

The MXB dictionary configuration in FLEx

Thomas Willett, SIL Mexico

The hybrid dictionary configuration in FLEx (FieldWorks Language Explorer, <https://software.sil.org/fieldworks/>) is based on the model of lexicography developed by the Mexico Branch of SIL and applied in an extensive series of published bilingual vocabularies and dictionaries in Mexican indigenous languages (www.mexico.sil.org/publications/i-vimsa), as well as for languages of other countries. The theoretical basis for this model is explained by Doris Bartholomew and Louise Schoenhals in *Bilingual Dictionaries for Indigenous Languages* (*BDIL*, www.mexico.sil.org/es/resources/archives/2384).

Focus on the native speaker

Unlike most models of lexicography, the MXB model is focused on the native speakers of indigenous languages as the primary users. Given that for many of them, the bilingual dictionaries we make will be the first ones produced in their languages, and that reading their languages is a relatively new experience, this model keeps the presentation of dictionary entries *as simple as possible*, especially at the beginning. Thus, for example, etymological information and lists of inflectional variants, such as plurals and past participles are presented *after the basic information* for each sense is presented, instead of after the headword, as is the common practice in dictionaries of international languages.

Specifically, each dictionary article has the following general structure (*BDIL*, 25-26), where the field names in FLEx are given in bold.

1. Entry form in the vernacular:
 - a. **Citation form**, an actual word written in the practical orthography;
 - b. **Pronunciation** information (optional), e.g. tone information not included in the practical orthography.
2. Sense discriminations translated into the language of wider communication:
 - a. Grammatical category (**Grammatical information**) usually abbreviated;
 - b. Translation equivalent (**Definition**) of the entry form in the language of wider communication. If there is no equivalent form, a clear and concise explanation (**Definición descriptiva** = Descriptive definition) of its usage is given, e.g. for grammatical morphemes;
 - c. Qualifying comment (**Restrictions**) on the Definition, where necessary, to keep the translation equivalent (Definition) as short as possible;
 - d. Illustrative sentence (**Example**) that demonstrates the use of the entry word in a natural context;
 - e. Natural free **Translation** of the sentence in the language of wider communication.

These two types of information are essential to every entry. That is, the entry word and its senses make up the *basic entry*. All the other information that can be included is optional and amplifies the information included in the basic entry.

3. Minimal linguistic information of three types:
 - a. The components of complex entries such as derived forms and compounds;
 - b. The conjugation class or “principal parts” of a paradigm (**Inflectional Variants**);
 - c. A grammatical clarification (**Note**) necessary for correct usage of the word;
4. Cross references of four types:
 - a. **Pronunciation Variants** (*Var.*);
 - b. **Regional** (dialectal) **Variants**, preceded by an abbreviation of the region where the variant is used;
 - c. **Synonyms** or **Antonyms**, which are usually main entries;
 - d. **References** (*Véase*) to other words that are morphologically related.
5. Subentries which refer to other entries that share the same entry word:
 - a. **Derivatives** (e.g. nouns from verbs, transitive verbs from intransitive verbs);
 - b. **Compound** words utilizing the same root;
 - c. **Idioms** (idiomatic phrases) that include the entry word.

Native speakers may not need or want to refer to the optional information provided by the last three types, but linguistically aware speakers and non-speakers, especially linguists, may find this information useful.

In order to help the reader keep track of these different types of information, the MXB hybrid configuration maintains a distinction between citations from the *indigenous* language, which are always shown in boldface type; translations into the *analysis* language, which are always shown in regular type; and *metadata* or technical information, which is always shown in italics. In addition, colors are used in FLEx to help the compiler further distinguish between various types of information in the entry: Indigenous language citations are in blue, translations are in green, and metalanguage is in yellow (when referring to meaning) or red (when referring to grammatical form).

Meaning in the basic entry

Of the information provided in the basic part of the entry, the **Citation form**, the **Definition**, and optional **Restrictions** work together to provide a clear and concise presentation of the meaning of each lexeme.

Citation form

The focus on the native speaker begins by citing a form for each lexeme that is recognized as legitimate in a spoken utterance, rather than an abstract form from which it can be derived. For example, verbs are cited in the form closest to the “infinitive” in the language of wider communication, which could be the future or present tense for third person singular.

This increases the chances that speakers will find the dictionary of interest to them, which in turn serves as an aid to literacy. Furthermore, with more and more native speakers of indigenous

languages compiling their own dictionaries, citing a naturally spoken form for each lexeme is much easier for them to manage than are abstract forms, especially in the early stages of analysis.

However, it is often helpful to maintain a distinction between the **Citation form**, which is an actual spoken word, and the **Lexeme form**, which is an underlying or abstract root of the entry word. When this is done, the Lexeme Form and the **Gloss** are used for text analysis, while the Citation Form and the **Definition** are used for lexical analysis.

Definitions and glosses

Given that a bilingual dictionary does not have “definitions” such as those found in monolingual dictionaries of international languages, but rather *translation equivalents* in the language of wider communication, the MXB configuration makes a distinction between a **Definition** and a **Gloss**.

One invariant *gloss* is ordinarily assigned to each root morpheme, both lexical and grammatical, as the “label” for that morpheme in the analysis language. These glosses are primarily used in glossing texts so that the analyst can keep track of homophonous morphemes and others which can easily be confused.

In contrast, one or more *translation equivalents* in each analysis language are normally given as the *definition* of each sense of each lexeme in a bilingual dictionary. Focus on the native speaker is furthered by citing the *equivalent* word or words in the language of wider communication as the **Definition** of each sense of a lexeme. This teaches the user what is the most appropriate way to translate it when speaking the language of wider communication.

Keeping distinct translation equivalents for each sense also makes it easier for the compiler to create the *Reversal index* in the language of wider communication. When these “definitions” are short, they are easily converted into headwords in the Reversal index, which would be much harder if the definitions were long and complicated, as in monolingual dictionaries.

When there is no translation equivalent in the language of wider communication, the MXB configuration provides a custom field called **Definición descriptiva** (Spanish for “descriptive definition”). This field is used *instead of a Definition*. It is often useful for describing the meaning of grammatical morphemes. For example, the English suffix *-graphy* could be shown in following way in an English-Spanish dictionary:

-graphy *suf. s. Indica la acción de escribir; p. ej. lexicography* lexicografía; **typography** tipografía.

This descriptive definition (in English: *Indicates the action of writing; e. g. ...*) is normally followed by two short examples of the use of the affix in the given sense. Together, this descriptive definition plus the examples are punctuated as one grammatical sentence. This, plus the fact that the Descriptive definition occurs in italics, sets it off as clearly different from a translation equivalent.

Qualifying comments

Since translation equivalents between languages seldom share all the same senses, qualifying comments are often needed in the **Restrictions** field. These comments restrict the sense of the translation equivalent to the corresponding one in the indigenous language. For example, many languages have multiple words for “carry” depending on how the object is carried or what part of

the body is used to carry it. In these cases the dictionary compiler would translate each one as “carry” with the appropriate restriction, such as “on one’s back” or “in one’s hands.”

The fact that these restrictions are in a separate field means they can also be used to qualify the indigenous lexemes in the Reversal index. For example, on the English side of an English-Spanish dictionary database, the entry for “entry” could have one sense qualified as “in a dictionary.”

entry *n.* entrada (*en un diccionario*)

Then, in the Spanish reversal index, the corresponding entry could use the same qualification on the English word (where *f.* stands for a feminine noun).

entrada *f.* entry (*en un diccionario*)

The main difference between these two entries is which language comes first, but the restriction on the meaning of the sense in question is the same for both.

Emphasis on illustrative sentences

Another major distinctive of the MXB model is that it makes illustrative sentences a primary focus of the dictionary article (*BDIL*, xv). Illustrative sentences serve both the compiler of the bilingual dictionary and the user.

During the process of eliciting illustrative sentences, the compiler becomes aware of sense discriminations, co-occurrence restrictions on classes of lexical items, or grammatical restrictions which might otherwise have been overlooked (*BDIL*, 59).

For the user, the illustrative sentences serve more than one purpose. Native speakers will resonate with the naturalness of the examples and exclaim, “Yes, that’s the way we talk!” This provides a source of pride in their language and a means of demonstrating their unique form of expression to the outside world.

For non-speakers, the illustrative sentences provide clear examples of correct morpho-syntactic usage in the language. They also open a window onto the indigenous culture which amply illuminates their unique form of thinking. Cultural observations (entered into the **Encyclopedic information** field) can provide further explanations of customs mentioned in the sentences.

The rationale for the inclusion of illustrative sentences is explained further in chapter 9 of *BDIL*, where it is shown that they perform one or more of the following functions for each sense of each word:

- 9.1. They clearly delineate the area of meaning of the entry word.
- 9.2. They illustrate the use of the word in a correct grammatical context.
- 9.3. They distinguish between applicable and non-applicable senses and contexts of the gloss.
- 9.4. They clarify potential ambiguities due to multiple meanings of the translation equivalent.
- 9.5. They display the local culture and local literary style, providing a glimpse into the culture's unique worldview.

Other topics

Although the parts of *BDIL* that refer to computational applications of the MXB configuration are now archaic, many valuable discussions of lexicographical topics are included, such as:

- The lexical unit (chapter 1)
- Three levels of lexical analysis (chapter 2)
- The procedures of lexical analysis (chapter 3)
- The psychologically natural citation form (chapter 5)
- Practical criteria for selection of the citation form (chapter 6)
- The translation equivalent in the language of wider communication (chapter 8)
- Sense discriminations (chapter 10)
- Qualifying comments (chapter 11)
- Variant pronunciations and geographical variants (chapter 13)
- Translation equivalents and refinement techniques (chapter 14)
- The index in the language of wider communication (chapter 15)
- Loan words in the bilingual dictionary (chapter 16)
- Special grammatical designations for indigenous languages (chapter 18)
- Essential linguistic information and cross-referencing (chapter 19)
- Subentries (chapter 20)
- Evaluation of the dictionary article (chapter 26)

Conclusion

The MXB configuration focuses on native speakers as the primary users of the dictionary. It does this by presenting first, within each entry, the information most relevant to them; by making use of spoken forms as citation forms; by providing translation equivalents in the language of wider communication, which may be qualified by restrictions; and by employing grammatically and culturally natural illustrative sentences and translations to illustrate sense discriminations. Over the last several decades, this method has proven effective in introducing the first bilingual dictionary into many indigenous communities in Mexico and elsewhere.

Appendix 1

The following examples of each type of entry (Main, Minor, and Complex Form) are taken from the “MXB modelo” project in which English is the vernacular language and Spanish is the analysis language. The configured format from the Dictionary view is shown first, then what fields are used in Lexicon Edit to properly enter the displayed information. A unique feature of this hybrid configuration is that Complex Forms can appear *both as main entries and as subentries*.

Main Entry

entry *s.* *entrada* (en un diccionario o vocabulario) **Many dictionaries contain entries for words that are no longer in use.** Muchos diccionarios contienen entradas para palabras que ya no se usan. (*La costumbre de incluir palabras obsoletas es antigua.*)
 [pl. **entries**] *Sinón.* **article**
 ♦ **entry word** *s.* *palabra de entrada* (en un diccionario o vocabulario)

Lexeme Form	entry
Morph Type	root
Citation Form	entry
Dialect labels (Entry)	
Complex Forms	entry word
Components	
Variant of	
Pronunciation	
Note	
Restrictions	
Summary Definition	
Cross References	
Synonym	article
Descripción Breve(*)	

□Sense 1	
Reversal Entries	<i>entrada</i>
Definition	<i>entrada</i>
Restrictions	<i>en un diccionario</i>
Grammatical Info.	sustantivo ▼
Example	Many dictionaries contain entries for words that are no longer in use.
Translation	<i>Muchos diccionarios contienen entradas para palabras que ya no se usan.</i>
Encyclopedic Info.	<i>La costumbre de incluir palabras obsoletas es antigua.</i>
Definición descriptiva(*)	
Comentario restrictivo(*)	<i>en un diccionario o vocabulario</i>

□Variants	
Variant Form	entries
Variant Type	<i>plural</i>

□Publication Settings	
Publish Entry in	<i>Main Dictionary</i>
Show As Headword in	<i>Main Dictionary</i>
Subentries	entry word

An asterisk (*) indicates a custom field with a Spanish name. **Descripción breve** [“Brief Description”] is the entry level equivalent of **Definición descriptiva** [“Descriptive Definition”]; it is used for minor entries (as in the sample entry for **-graphy** shown on page 3).

The cited **Headword** is always the **Citation Form**, if it exists; otherwise, it is the **Lexeme Form**. If the Lexeme Form is used for root forms to facilitate text glossing, the compiler should be sure an appropriate Citation Form is also filled in for the dictionary headword.

The **Pronunciation** field is used to specify phonological details not evident in the practical orthography. It can take the form of a phonetic transcription (as shown below) or as tone patterns, such as: a-b.m.b.

The **Note** field is used for a grammatical or etymological observation (as shown below), *not* for semantic distinctions, which should be made in the Restrictions field.

Inflectional **Variants** and Pronunciation variants create separate, minor entries which need to be completed with a Citation Form and a Summary Definition (as shown below).

The **Publication Settings** fields are used to include or exclude entries from a publication (**Publish Entry in**) or show or hide them as entries in a publication (**Show As Headword in**). Subentries can be manually sorted using the **Subentries** field.

Minor Entry (Variant)

entries [*'en.triz*] {*pl. de entry*} *entradas (en un diccionario o vocabulario)*

Lexeme Form	entries
Morph Type	stem
Citation Form	entries
Dialect labels (Entry)	
Complex Forms	
Variant Type	plural
Variant of	entry
Components	
Pronunciation	<i>'en.triz</i>
Note	
Restrictions	<i>en un diccionario o vocabulario</i>
Summary Definition	<i>entradas</i>
Cross References	
Descripción Breve(*)	
<input type="checkbox"/> Publication Settings	
Publish Entry in	<i>Main Dictionary</i>
Show As Headword in	<i>Main Dictionary</i>

Minor entries are created in the Variants section of a Main Entry by filling in the **Lexeme Form** and **Variant Type**. The headword of the Main Entry becomes the **Variant of** field of the Minor Entry.

All Variants that are cited in another entry of a Publication must be included in the **Publish Entry in** for that Publication. However, they do not have to be shown as headwords in that Publication.

If the Minor entry shows as a Headword, the rest of the Minor entry can be manually edited to include a **Citation Form**, any **Dialect labels** or **Pronunciation** desired, and a **Summary Definition** (with optional **Restrictions**) or optionally, a **Descripción Breve**.

Main Entry (Complex Form)

A Complex Form Entry is a Main Entry that has **Components**. When the first component is entered, the **Component Form Type** field shows up. In addition, the **Show Subentry under** field appears in the Publications Settings. The first component is shown there by default, but the user may decide whether to show the Complex Form as a subentry under all, only some, or none of its components.

All the information for subentries comes from the corresponding Complex Form entry: (a) the subentry form comes from the **Citation Form**; (b) the part of speech comes from the **Grammatical Info.** field(s) in the sense(s); (c) the definition comes from the **Summary Definition** field; and (d) the qualifying comment comes from the **Restrictions** field (the one immediately above Summary Definition, *not* the one in the senses).

Since the **Restrictions** field automatically serves as the qualifying comment for the vernacular word in the Reversal Index, the custom field **Comentario restrictivo** (“Restrictive Comment”) allows users to have a different qualifying comment on the vernacular side, as in the above examples. Otherwise, this field can be eliminated from view and the Restrictions field can be configured to serve as the qualifying comment on the vernacular side, too.

entry word *s. palabra de entrada (en un diccionario o vocabulario)* **The entry word of a dictionary article is usually written in the practical orthography. Normalmente la palabra de entrada del artículo de un diccionario se escribe con la ortografía práctica.**

Se pronuncia con intonación de frase. Véase entry, word

Lexeme Form	
Morph Type	phrase
Citation Form	entry word
Dialect labels (Entry)	
Complex Forms	
Complex Form Type	Compound
Components	entry word
Variant of	
Pronunciation	
Note	Se pronuncia con intonación de frase.
Restrictions	en un diccionario o vocabulario
Summary Definition	palabra de entrada
Cross References	
Descripción Breve*	

☐Sense 1	
Reversal Entries	entrada: palabra de entrada palabra: palabra de entrada
Definition	palabra de entrada
Restrictions	en un diccionario
Grammatical Info.	sustantivo ▼
Example	The entry word of a dictionary article is usually written in the practical orthography.
Translation	Normalmente la palabra de entrada del artículo de un diccionario se escribe en la ortografía práctica.
Encyclopedia information	
Definición descriptiva(*)	
Comentario restrictivo(*)	en un diccionario o vocabulario

☐Publication Settings	
Publish Entry in	Main Dictionary
Show As Headword in	Main Dictionary
Show Subentry under	word

Reversal index

The **Headword** for a Reversal Entry must be entered in the **Reversal Entries** field. In order to respect Spanish lexicographic tradition, the Headword should always be a single orthographic word. All Headwords from the same sense are entered in this line, each followed by a faint line that appears when a new word is entered.

When the Definition is a phrase, it should occur as a **Reversal Subentry** under one or more Headwords. This is also indicated in the Reversal Entries field by this use of a colon: the word before the colon will become the Headword in the Reversal Index, and the words after the colon will become a Reversal Subentry under that Headword.

The **Restrictions** field automatically serves as the qualifying comment for the vernacular word in the Reversal Index.

The Reversal Index must be manually edited to include the appropriate parts of speech in Spanish.

The following Reversal Index entries show what the Reversal Entries fields in the above entries would produce, with the appropriate Spanish category added.

...

entrada *f.* entry (*en un diccionario*)

◆ **palabra de entrada** entry word (*en un diccionario*)

...

palabra *f.*

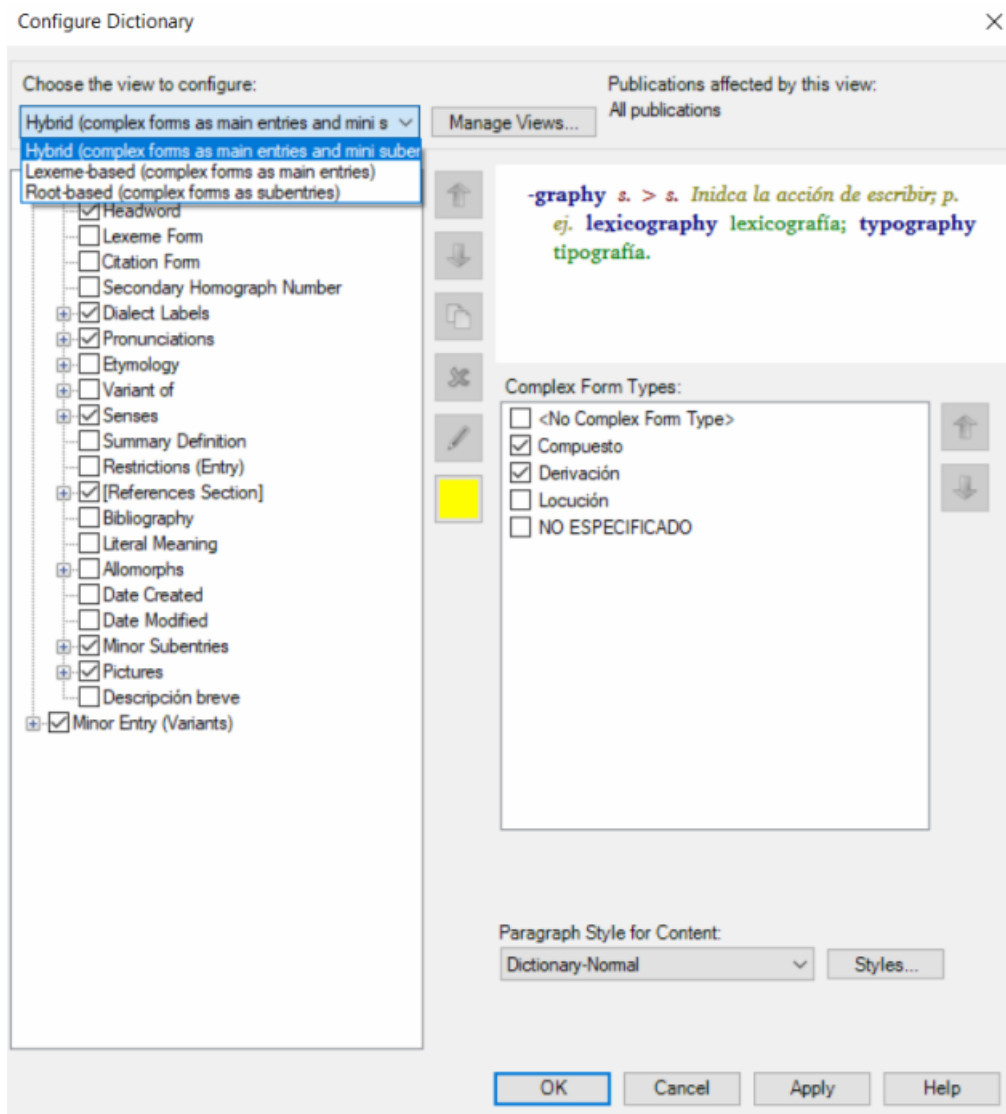
◆ **palabra de entrada** entry word (*en un diccionario*)

Appendix 2

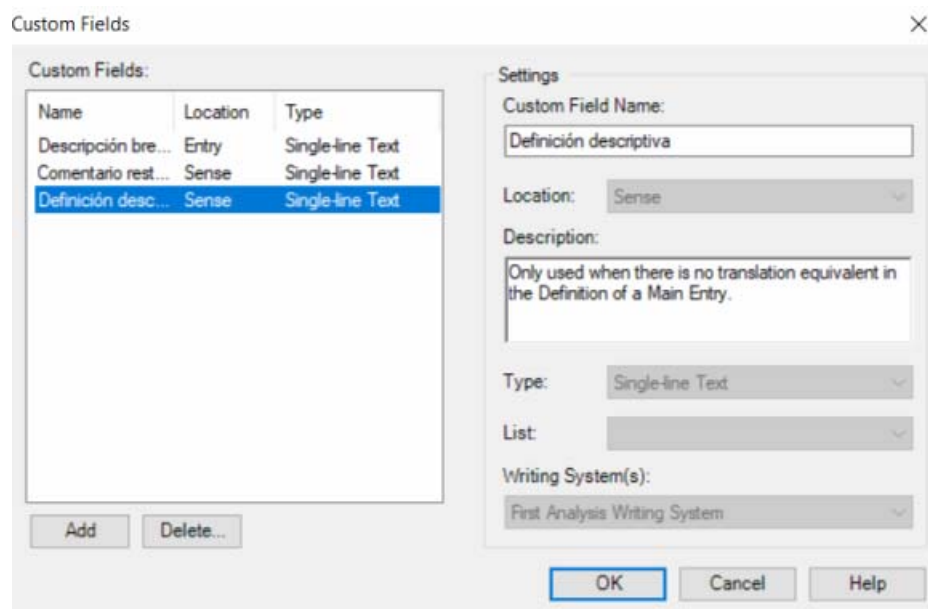
Although FLEEx allows the user to export a configuration from one project and import it into another, this process changes many other things, including the Grammatical Categories, some of the Styles, and some of the lists, such as Complex Form Type. For this reason we do not recommend exporting and importing a configuration at this time (FW8.3.10).

Instead, we currently recommend using the Hybrid configuration that now comes with FLEEx as a third option, besides Lexeme-based and Root-based. It is based on the MXB configuration, but not all the details are the same. The following specifications are necessary to modify the standard Hybrid configuration to conform to our MXB norms for research and publication. For best results, they should be applied in the order indicated.

First, use Tools > Configure > Dictionary to choose the Hybrid configuration:



Second, in Lexicon Edit, use Tools > Configure > Custom Fields to create the following three custom fields:

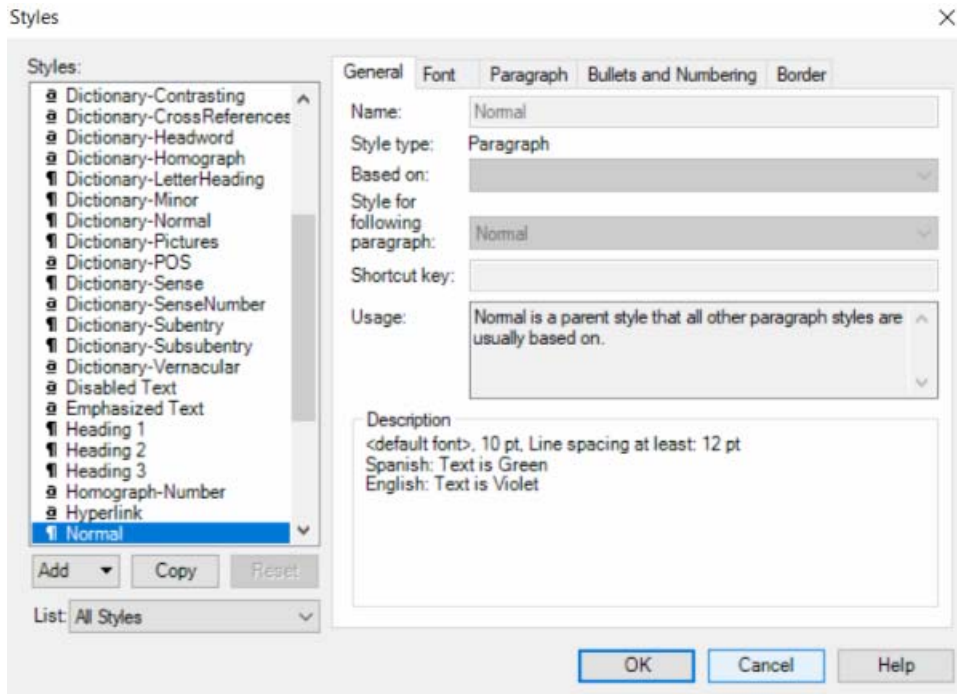


These fields all have Spanish names (including accents) and the following specifications:

Custom Field Name	Location	Description	Type	List	Writing System(s)
Descripción breve	Entry	Only used when there is no translation equivalent in the Summary Definition of a <i>Minor</i> Entry.	Single-line Text	—	(*)
Definición descriptiva	Sense	Only used when there is no translation equivalent in the Definition of a <i>Main</i> Entry.	Single-line Text	—	(*)
Comentario restrictivo	Sense	Qualifying comment to restrict the meaning in the Definition field of a <i>Main</i> Entry.	Single-line Text	—	(*)

(*) Specified in the Configuration.

Third, use Format > Styles to be sure they are correct:



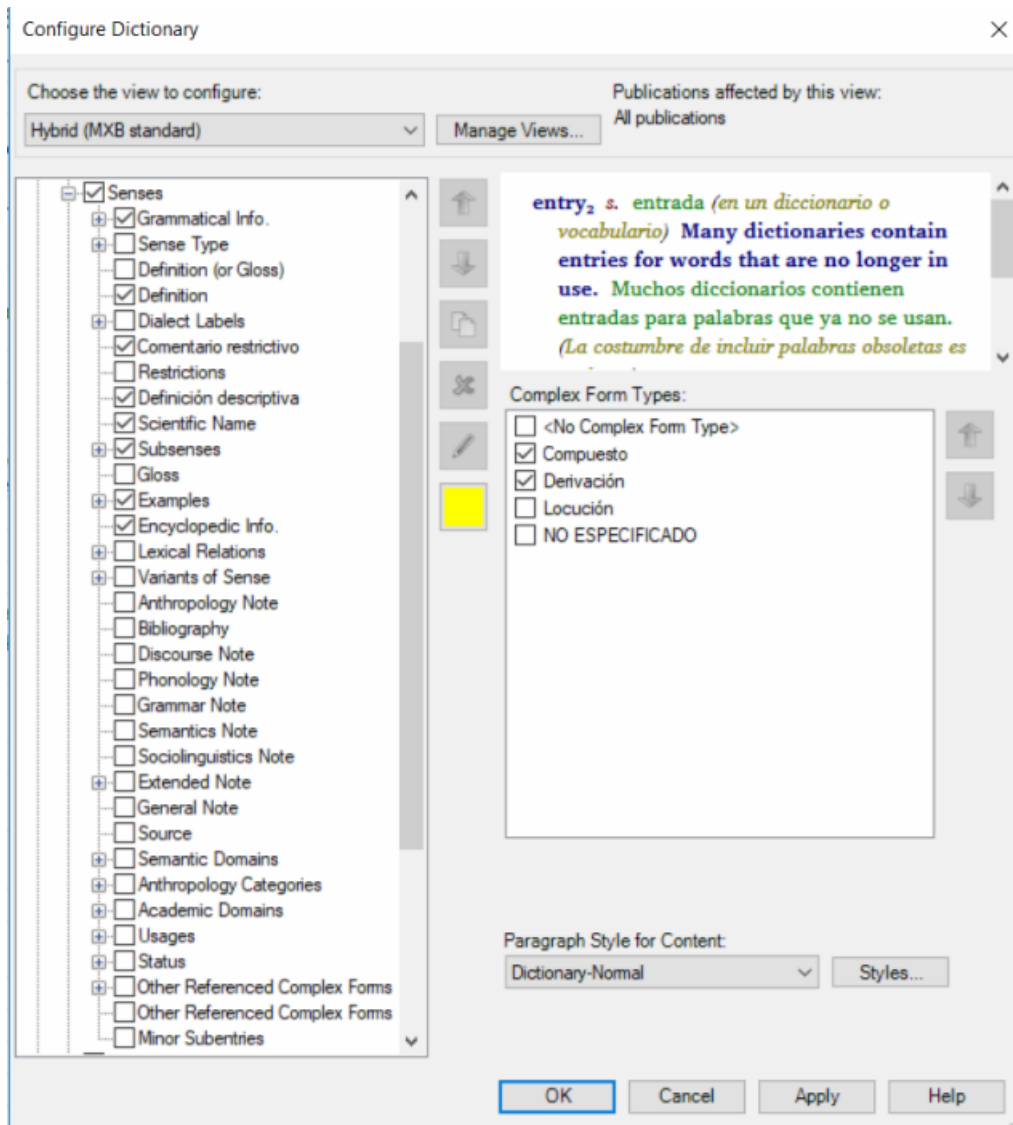
In the Font tab, click on each of the following styles and modify the information as specified (where “—” indicates neither checked nor blank):

Type	Style	Size	Color	Bold	Italic	Subscript
<u>a</u>	Dictionary-Contrasting	10	Dark Yellow	no	yes	—
<u>a</u>	Dictionary-CrossReferences	10	Blue	yes	no	—
<u>a</u>	Dictionary-Headword	11	Blue	yes	no	—
¶	Dictionary-Normal	10	Black	—	—	—
¶	Dictionary-Pictures	10	Black	—	—	—
<u>a</u>	Dictionary-POS	10	Dark Red	no	yes	—
¶	Dictionary-Sense	10	Black	—	—	—
<u>a</u>	Dictionary-SenseNumber	—	Black	yes	no	—
¶	Dictionary-Subentry	10	Black	—	—	—
<u>a</u>	Dictionary-Vernacular	10	Dark Blue	yes	—	—
<u>a</u>	Emphasized Text	10	Spanish: Dark Yellow English: Plum vernacular: Dark Blue	no no yes	yes no no	— — —
¶	Homograph-Number	—	Blue	yes	no	yes
¶	Normal	10	Spanish: Green English: Plum vernacular: Dark Blue	— — —	— — —	— — —
<u>a</u>	Reversal-Headword	11	Spanish: Green English: Plum	— —	— —	— —
¶	Reversal-Normal	10	Black	—	—	—
¶	Reversal-Sense	10	Black	—	—	—
¶	Reversal-Subentry	10	Black	—	—	—
<u>a</u>	Reversal-Vernacular	10	Dark Blue	yes	no	—

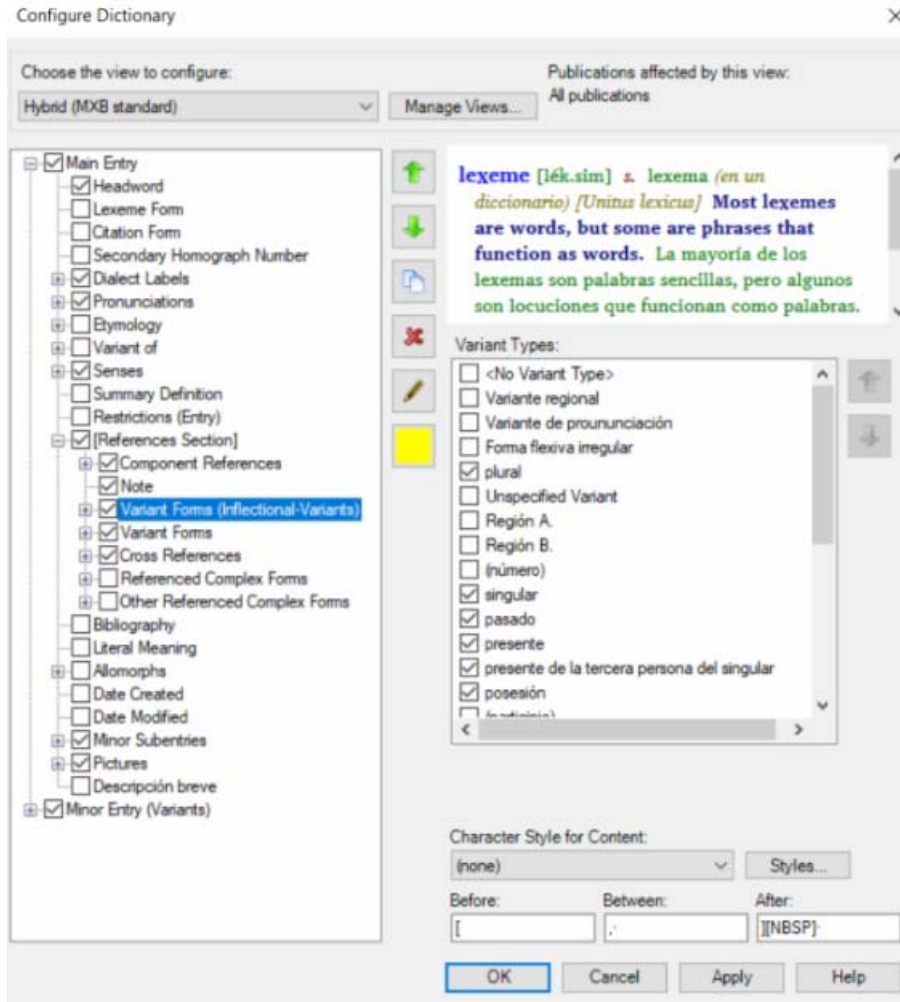
If one of the styles in this list does not exist, you can create it with the Add or Copy button.

Fourth, use Tools > Configure > Dictionary to configure the individual fields in both **Main** entries and **Minor** entries as follows.

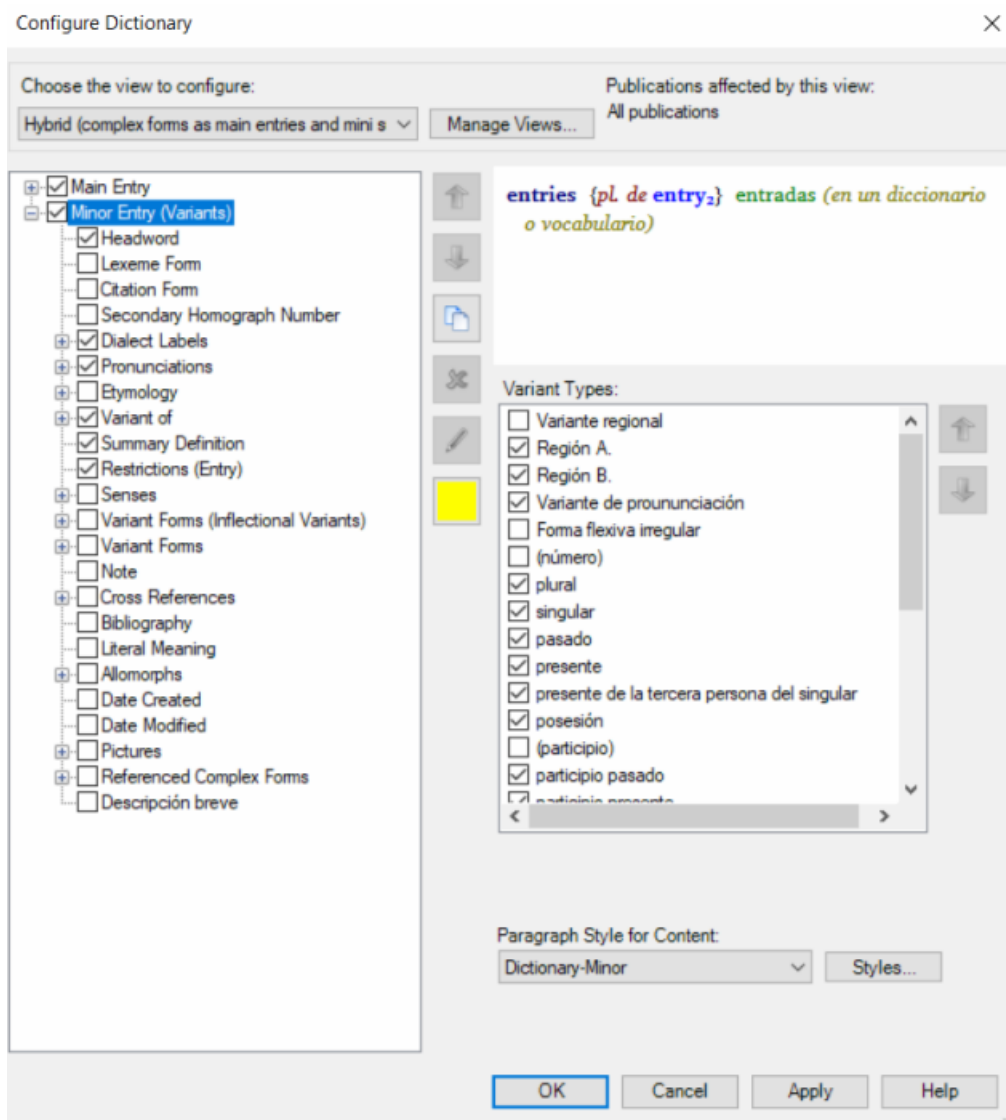
For *Main entries*, check for the appropriate fields in *Senses*:



Also for *Main entries*, check for the appropriate fields in the *References Section*:



For *Minor entries*, check for the appropriate fields as follows:



Summary

Surrounding text

Field	Before	Between	After
Entry Paragraph			
Headword			fixed space(*)
Pronunciation	“[”		”], fixed space
Sense Paragraph			
Sense number	space		“.”
Category	space		fixed space
Definition	space		fixed space
Comment			fixed space
Example	space		fixed space
Translation	space		fixed space
Encyc. Inf.	space		
Reference Paragraph			
Component ref.			fixed space
Note	space		fixed space
Inflectional var. space, “[”		“,”	”], fixed space
Cross ref.	space		fixed space

(*) Shows as as [NBSP] in FLEx.

Spacing

- a. There should be two spaces between each “segment” of the entry: Headword; Category; Definition; Example; Translation; References.
- b. There should be one space between “parts” of segments: Headword & Pronunciation; Senses Number & Category (or Definition); Definition & Restriction; Variant Label & Variant Form; Reference Label & Reference Form.

Senses paragraphs

- a. For entries with multiple senses, the first sense should follow the definition without a paragraph break. All subsequent senses should start with a paragraph break.
- b. Subsenses should be concatenated and joined by a comma and space.

Reference paragraph

- a. Component references with one space between Complex Form Type ("*Comp. de*" or "*Deriv. de*") and Referenced Headword.
- b. Note (free form grammatical observation, punctuated as complete sentence)
- c. Inflectional variants: [*abr1*, **form1**; *abr2*. **form2**]
- d. Cross references ("*Sinón.*", "*Antón.*" or "*Véase* plus **form**)
- e. Pronunciation variants (same dialect; "*Var.*" plus **form**)
- f. Regional variants (different dialects; "*Dial. Q.*" plus **form**)

Subentries

Configure as separate, bulleted paragraphs, further indented than the entry paragraph and with hanging indent.

Examples

The following excerpt is from the *Diccionario mixteco de San Andrés Yutatío, Tezoatlán, Oaxaca*. [Serie de vocabularios y diccionarios indígenas “Mariano Silva y Aceves” 49].