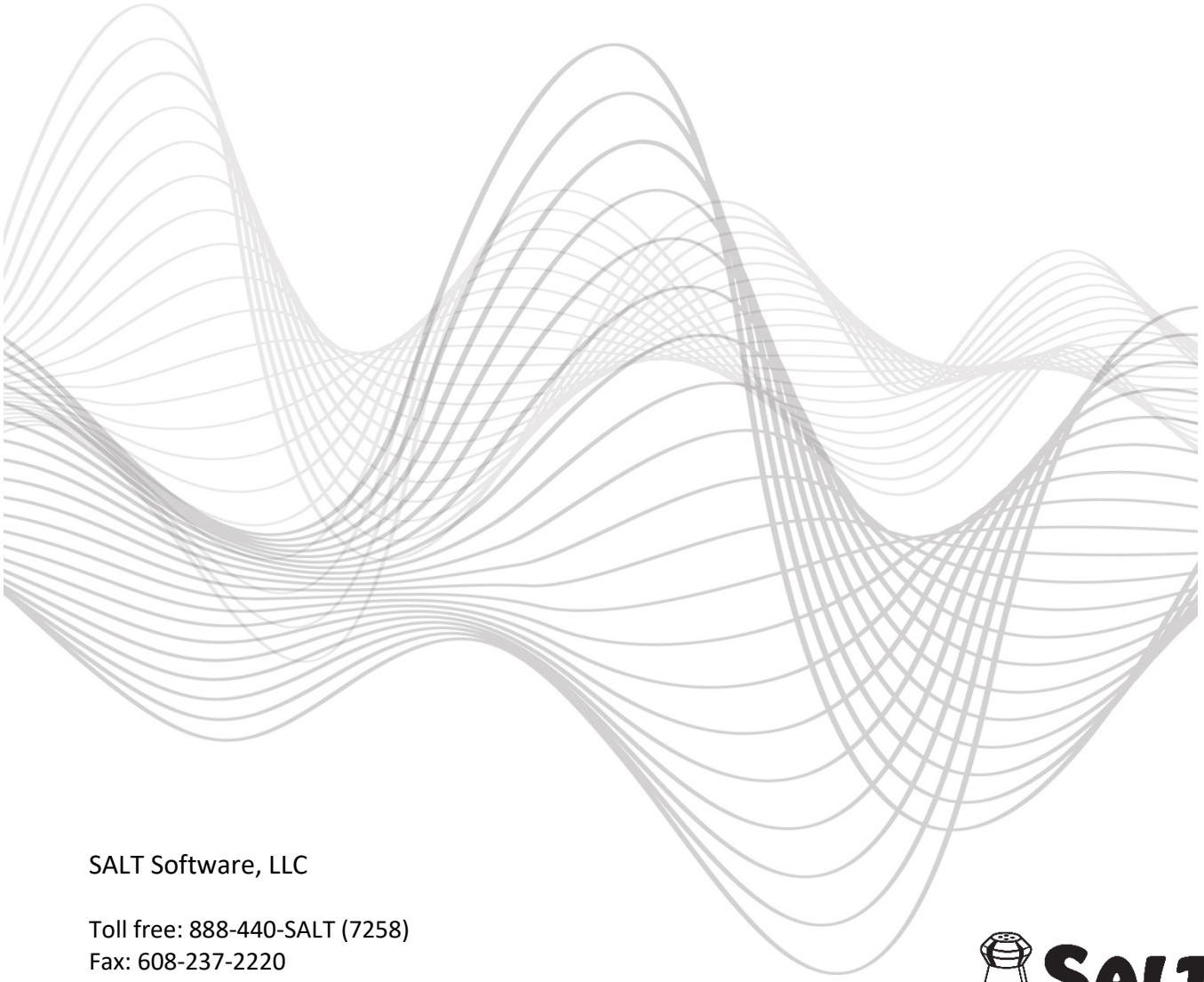
TRANSCRIPTION CONVENTIONS

Adapted for Spanish



SALT Software, LLC

Toll free: 888-440-SALT (7258)
Fax: 608-237-2220
support@saltsoftware.com
www.saltsoftware.com

Revision date: September, 2017

Acknowledgements

The implementation of Spanish transcription and analysis using SALT is the result of collaboration with Aquiles Iglesias and Raúl Rojas from Temple University. We would like to express our appreciation for their insights into the many issues involved in analyzing Spanish transcripts and their willingness to work with us to develop transcription conventions and analyses designed specifically for Spanish and bilingual transcripts. The Spanish root identification file, identifying the infinitive form for over 465,000 Spanish words, was derived from a reverse conjugation file which was licensed to us without charge by David German from the University of Victoria. Raúl Rojas did most of the editing needed to adapt this list for use within SALT. *The work at Temple University has been partially funded by research contract N01-DC-8-2100 from the National Institute on Deafness and Other Communication Disorders, National Institute of Health.*

Table of Contents

A. Using SALT with Spanish Samples

1. Overview

Our bilingual Spanish/English research project (Miller, Heilmann, Nockerts, Iglesias, Fabiano, & Francis, 2006) presented us with a major challenge; how to make consistent transcription decisions for Spanish that would allow comparison with English transcripts. In order to compare an individual's Spanish language skills with their English language skills we needed to be sure we were counting the same elements, words, morphemes, and utterances. Aquiles Iglesias and Raúl Rojas at Temple University collaborated with us to design the Spanish transcription format for SALT. It took the better part of a year to work out how to code specific features of Spanish which are inherently different than English, such as verb inflections and bound versus unbound clitics.

Most of the standard SALT transcription conventions can be applied to languages other than English. When adapting the transcription conventions to other languages, however, it is important to consider the structure and use of the language. It isn't enough to just write the words as they were spoken. You need to make sure that the representation you use results in an accurate analysis of the language. If using SALT to assess the language of bilingual speakers, it's important to develop transcription conventions to allow for comparison of the two languages.

Begin by defining how words are represented. After that, the rest should fall into place. For English samples, we define words by marking bound morphemes so that "look", "look/ed", and "look/ing" are all the same word. In inflected languages, such as French and Spanish, we identify the root form of the word using the vertical bar, e.g., "buscan|buscar", "buscando|buscar". We mark prefixes in prefix-based languages, such as Farsi (Persian).

2. Character Sets

SALT supports the UTF-8 character set which includes all the Spanish characters.

Are these characters built into the SALT program? No, other character sets are not built into the SALT program although they are accessible from within SALT in the following ways:

- **Using the "Edit menu → Insert Symbol" option in SALT.**

  The "Edit menu --> Insert Symbol" option in SALT allows you to set up a list of characters which can be inserted into the editor using minimal keystrokes. By default, the list contains the non-standard keyboard characters used for Spanish and French but this list can be edited to suit your needs.

- **<u>Windows OS</u> methods of accessing the diacritic characters**

  Method 1: Using the standard Windows® keyboard.

  This is an easy method but may or may not work for you, depending on your operating system and/or keyboard.

Hold down the CTRL key and type *(then release the keys)*	**Then type**	**To get**
' (single quote)	a, A, e, E, i, I, o, O, u, or U	á, Á, é, É, í, Í, ó, Ó, ú, or Ú
` (back quote)	a, A, e, E, i, I, o, O, u, or U	à, À, è, È, ì, Ì, ò, Ò, ù, or Ù
: (colon)	a, A, e, E, i, I, o, O, u, U, y or Y	ä, Ä, ë, Ë, ï, Ï, ö, Ö, ü, Ü, ÿ, or Ÿ
^ (caret)	a, A, e, E, i, I, o, O, u, or U	â, Â, ê, Ê, î, Î, ô, Ô, û, or Û

~ (tilde)	a, A, n, N, o, or O	ã, Ã, ñ, Ñ, õ, or Õ
, (comma)	c or C	ç or Ç

Other character combinations are also available.

Method 2: Add the "United States International" input language.

This option involves adding an input language to your keyboard. It's a one-time setup and allows you to switch between your default keyboard and the United States International keyboard. Rather than provide instructions for each operating system, use the help option to search for instructions.

Search for "add or change input language". Follow directions to add the "United States International" keyboard located within the "English (United States)" keyboard.

Once the keyboard is added, there should be an icon added to your system tray which looks like a small keyboard. This icon allows you to switch between your default keyboard and the US International keyboard. When the language setting is set for "United States - International" you can access most of the special characters as follows:

Hold down the RIGHT ALT key and type	**To get**		**First type " (double quote), then type**	**To get**
a or A	á or Á		u or U	ü or Ü
e or E	é or É		SPACE BAR	"
i or I	í or Í		Note: to type the double quote in your transcript, type the double quote followed by a blank space.	
o or O	ó or Ó			
u or U	ú or Ú			
n or N	ñ or Ñ			
number “1” key	¡			
slash key “/”	¿			

Other character combinations are also available.

- **3rd party software.**

  There are several free software applications which may be downloaded to provide easy access to the other character sets, e.g., www.onehourprogramming.com.

- **Use other word processor.**

  Alternatively, you may prefer to type the entire transcript in Microsoft Word or some other word processor where other character sets are available. Then, save the transcript as a text file with the extension .SLT. Or, cut/copy and paste the text from Microsoft Word (or other word processor) into the SALT transcript window.

B. Transcription Conventions Adapted for Spanish Samples

The implementation of Spanish transcription and analysis using SALT is the result of collaboration with Aquiles Iglesias from Temple University and Raúl Rojas from University of Texas at Dallas. The samples stored in the *Bilingual Spanish Story Retell* and *Bilingual Spanish Unique Story* databases were transcribed using these conventions.

1. Diacritic Characters

The use of diacritics can mark grammatical differences between identical word forms, therefore potentially altering the meaning of the word and utterance. Including them will make the transcript more valid. The use of an accent can affect how a word gets counted as part of a certain word list (e.g., in Spanish, “el” as the direct article meaning “the” vs. “él” the personal pronoun meaning “he”).

The characters "¿" and "¡" may be entered at the beginning of questions and exclamations. They are considered punctuation marks and are ignored for analysis. For example:

C ¿Rana dónde estás?
C ¡Vamonos, vamonos!

2. Utterance Segmentation into MC-units

Because Spanish is a pronoun-drop language (Bedore, 1999, 2001; Rojas & Iglesias, 2006), the transcripts in the Bilingual Spanish/English Story Retell databases were segmented into MC-units which were developed specifically for these samples to account for these omitted pronouns and to provide consistency in the transcription of both Spanish and English samples. Utterances containing successions of verbs without subjects are segmented and a fragment [F] code is placed at the end of each utterance lacking a stated subject as a result of this segmentation.

3. Root Identification

Root identification instructs SALT to consider the word immediately preceding the "|" symbol as the word which was actually said, and the word immediately following to be the word root. This convention was originally created to identify Spanish verb forms but has been expanded to identify Spanish diminutives and Spanish superlatives, as well as words repeated multiple times for emphasis and overgeneralization errors. Routines have been built into the SALT software to automate this identification process.

a. Spanish Verbs

The highly inflected morphology of Spanish can significantly affect the post-inflected root word/stem. The word root identification convention, vertical bar "|", is used to credit Spanish-speakers for exhibiting use of morphological forms as well as to avoid over-inflation of the number of different words (NDW) used. If a speaker produces a variety of inflected forms of the same word within a transcript (e.g., es, son, eran, éramos), each production is coded to identify the root word (e.g., es|ser, son|ser, eran|ser, éramos|ser). The speaker would be given credit morphologically for producing different words, but each inflected form would be considered an inflected variation of the same root word, “ser”.

For example:

C **Había|haber** una vez un niño que **tenía|tener** una rana.

Root identification instructs SALT to consider the word immediately preceding the "|" symbol as the word which was actually said (i.e., "había" and "tenía"), and the word immediately following to be the root word (i.e., "haber" and "tener").

<u>Auxiliary Verbs</u>

When a verb is used as an auxiliary, precede the root identification with "aux", (e.g., "auxestar",

"auxandar", "auxsalir"). For example:

1) C El perro **está|estar** con las abeja/s.
2) C El niño **estaba|auxestar** buscando|buscar la rana en el árbol.

In the first utterance, "está" is used as a main verb and identified as a form of "estar". In the second utterance, "estaba" is used as an auxiliary verb and identified as a form of "auxestar".

b. Spanish Diminutives

In Spanish, diminutives are bound morphemes that change the meaning of a word by indicating diminution. Diminutives in true form are suffixes attached to an object and are used in reference to something of a larger size. The following is a list of known diminutive suffixes: -ete, -eta, -ico, -ica, -ito, -ita, -illo, -illa, -uco, -uca, -ucho, -ucha, -uelo, -uela

The use of diminutives may be so pervasive in some Spanish dialects that the morphological inflection of the diminutive may no longer indicate diminution. In order to prevent potential mean length of utterance in morphemes (MLUm) inflation for speakers who use a high frequency of diminutives without strictly indicating diminution, it was decided that diminutives (i.e., perrito, casita) should not be counted as bound morphemes. Instead, diminutives are coded as main body words derived from the corresponding root words. Therefore, coding for diminutives follows the root identification convention.

For example:

C Él dice|decir **ranita|rana** dónde estás|estar?
C El **perrito|perro** tumbó|tumbar las abeja/s.

c. Spanish Superlatives

Superlative forms, suffixes indicating that an object is larger than a referent, are not as pervasive in the Spanish language as are diminutives. Some common superlative suffixes include: -ote, -ota, -ísimo, -ísima. Like diminutives, coding for superlatives follows the root identification convention.

For example:

C Y el niño se[x] subió|subir en un árbol bien grandote|grande.
C Y el agua estaba|estar friísima|fría.

d. Automating Root Identification

To simplify transcription, lookup files containing words with their corresponding root forms are available. These root identification files (RIFs) are used to automatically identify a different word root than the one which was produced. The "Identify Roots" command in the Edit menu looks up all words in the transcript that have not been previously identified with the vertical bar. If the word is not found in the active RIFs, that word is ignored. If only one choice is found, the word is automatically identified. If a word contains more than one root option, the user is presented with a list of choices to select from. Note: the active RIFs are selected using the "Language Settings" option in the Setup menu. There are two Spanish RIFs:

Spanish Verbs contains a complete list of over 469,000 verbs. The only verbs intentionally left out of the file are “las”, “la”, “les”, “le”, “lo”, “una”, and “uno” due to their word form overlap with specific articles and pronouns. You should use this automation feature to identify all the verbs in your transcript.

Spanish Nouns and Clitics contains approximately 1,300 of the words used most often to retell the *Frog, Where Are You?* story. Because this is not a complete list, you should identify the plurals, diminutives, superlatives, and bound pronominal clitics as you are typing your transcript. This file may then be used to catch those you miss. *Refer to Section 4 for a discussion on marking bound morphemes and to Section 5 for a discussion on marking bound pronominal clitics.*

4. Plural Bound Morpheme

SALT uses bound morphemes to mark the use of plurals, possessives, verb inflections, and contractions. The same is not true for Spanish as possessives, verbs conjugations, and contractions operate under a distinct inflectional system which is often not amenable to the attachment of bound morphemes. Plurals are the only bound morphemes marked, e.g., “rana/s”.

5. Bound Pronominal Clitics

a. About Pronominal Clitics

As the name suggests, pronominal clitics are unstressed object pronouns which can be prosodically bound, morphosyntactically bound, or both. Pronominal clitics must occur with a verb, because they are verb-related as direct or indirect objects. They can be located preceding the verb as a *proclitic* ,e.g., lo buscó, or positioned after the verb as an *enclitic*, e.g., buscarlo. Pronominal clitics can move within an utterance, which changes the form but not necessarily the content of the utterance. Accordingly, these pronouns can occur as freestanding clitics or as bound clitics.

b. Why code for bound pronominal clitics?

Spanish has great word order flexibility. The Spanish language is not as dependent on word order for meaning as is the English language. In order to limit over-inflation of MLU and control the effects of dialectical variation, bound (not freestanding) pronominal clitics are identified with a plus sign ‘+’.

The freedom of movement that pronominal clitics possess is an important aspect in individual differences across language development and dialect. For example, a speaker could say “give it to me” in two ways:

1a) C me lo das. *Give it to me.*
1b) C dámelo. *Give it to me.*

Regardless of which utterance the speaker produces, (1a) or (1b), the content of the two utterances remains constant. Due to strict rules of spelling convention, (1a) is written as three separate words, and (1b) is written as one word. Herein lies the temptation to assign three words to (1a), and only one word to (1b). However, it is important to remember that SALT transcription analysis is based on oral language, not on written language. Therefore, both utterances should receive the same morphological credit.

Utterance (1b) should be coded for bound clitics in the following manner so that the word “dámelo” is counted as three words:

2b) C dá+me+lo. *Give it to me.*

The "+" symbol indicates the use of bound clitics by the respective personal pronouns. Whether or not pronouns indicate clitic-usage, they are still analyzed as separate main body words and separate

root words. Thus, utterances (1a) and (2b) will be given equal weight in the analysis. Both utterances possess the same verb and object pronouns; they have equal morphological value.

It should be clear that pronominal clitics can be bound or they can be freestanding, depending on the order of the pronoun(s) in relation to the verb. What is constant is that pronominal clitics always appear with a verb, even though they do not always stand immediately next to the verb.

For example:

C Él está **gritando+le** a la rana. *He is screaming at the frog.*
C Él **le** está **gritando** a la rana. *He is screaming at the frog.*

c. Bound morphemes versus bound pronominal clitics

Bound morphemes are marked with a slash, e.g., rana/s, and bound pronominal clitics are marked with a plus sign, e.g., buscar+lo. Bound morphemes receive morpheme credit but not word credit. Thus "rana/s" would be counted as one word with two morphemes. Bound clitics receive both word and morpheme credit. Thus "buscar+lo" would be counted as two words and two morphemes.

Omissions

Suppose the speaker said "él está gritándo a la rana". The omitted pronominal clitic would be coded as:

C Él está gritándo+*le a la rana.

or as

C Él *le está gritando a la rana.

In the first transcription the omission is treated as an omitted bound clitic. In the second transcription, the same omission is treated as an omitted unbound clitic.

6. Spelling Conventions

a. Yes words: OK, AHA, MHM, UHHUH, SÍ
b. No words: NO, AHAH, MHMH, UHUH
c. Filled pause words: AH, EH, ER, HM, HMM, UH, UM, and words coded as [FP].

7. Reflexive vs Non-reflexive Pronouns

Some pronouns can be used both reflexively and non-reflexively, with an overlap in word form (i.e., ME, TE, SE, OS, and NOS). The Spanish Standard Word Lists assume the reflexive personal pronouns are transcribed with the word code [X] and the non-reflexive pronouns are transcribed without the [X]. This difference is necessary to avoid giving a speaker credit for using both a reflexive and non-reflexive pronoun simultaneously (*since the pronoun words are identical*). If reflexive pronouns are not coded this way, they will be counted with the non-reflexive personal pronouns instead of with the reflexive personal pronouns. The word code, [X], is used to designate the use of reflexive pronouns in utterances such as:

C El niño **se[X]** fue|ir con el perro. *The boy left with the dog.*
C El niño dijo yo **me[X]** voy para la casa. *The boy said I go home..*

On the other hand, [X] is not used when the pronoun is not reflexive:

C El perro **me** ayudó a conseguir la rana. *The dog helped me find the frog.*
C El niño **se** la lleva a su casa. *The boy takes her to his home.*

Reflexive Verbs and Their Conjugations

Reflexives are used to reflect action done to oneself, himself, herself, etc, as noted in the examples below. Personal pronouns can be used reflexively or non-reflexively. Reflexive pronouns accompany reflexive verbs, and the third person reflexive pronoun "se" may occur in clauses where the subject is ambiguous (he, she, they).

To review, some reflexive verbs are exemplified:

Reflexive	**Translation**
me meto	I[myself] get into.
te levantas	You[yourself] get up.
nos acostamos	We[ourselves/each other] lie down.
se cae	He[himself]/she[herself]/it[itself]/you[yourself](formal) fall(s).
se caen	You[yourselves](formal)/they[themselves/each other] fall.
se baña	He[himself]/she[herself]/it[itself]/you[yourself](formal) bathe(s).
se bañan	You[yourselves](formal)/they[themselves/each other] bathe.

"Romance Reflexive" pronouns accompany intransitive verbs. Intransitive verbs do not require a direct object, but they may take one in certain environments (i.e., with romance reflexives). For example, "El perro se[X] cayó." *The dog fell.* In this example, "se" is a romance reflexive that accompanies the intransitive verb "cayó." "Cayó" is considered intransitive because it can stand on its own or may take the romance reflexive pronoun "se."

El perro cayó. or El perro se cayó.

References

Bedore, L.M. (1999). The acquisition of Spanish. In O. Taylor and L. Leonard (Eds.), *Language Acquisition Across North America: Cross-cultural and cross-linguistic perspectives* (pp. 157-208). San Diego, CA: Singular Publishing Group, Inc.

Bedore, L.M. (2001). Assessing morphosyntax in Spanish-speaking children. In A. Iglesias (Ed.), Communicative Assessment of the Hispanic Child. Seminars in Speech and Language, 22:1, 65-77.

Miller, Heilmann, Nockerts, Iglesias, and Fabiano (2006), "Oral Language and Reading in Bilingual Children", *Learning Disabilities Research & Practice*, 21(1), 30-43, 2006.

Rojas, R., & Iglesias, A. (2006). Bilingual (Spanish-English) narrative language analyses: Why and how? *Perspectives on Communication Disorders and Sciences in Culturally and Linguistically Diverse Populations*, *13:1*, 3-8.

Index