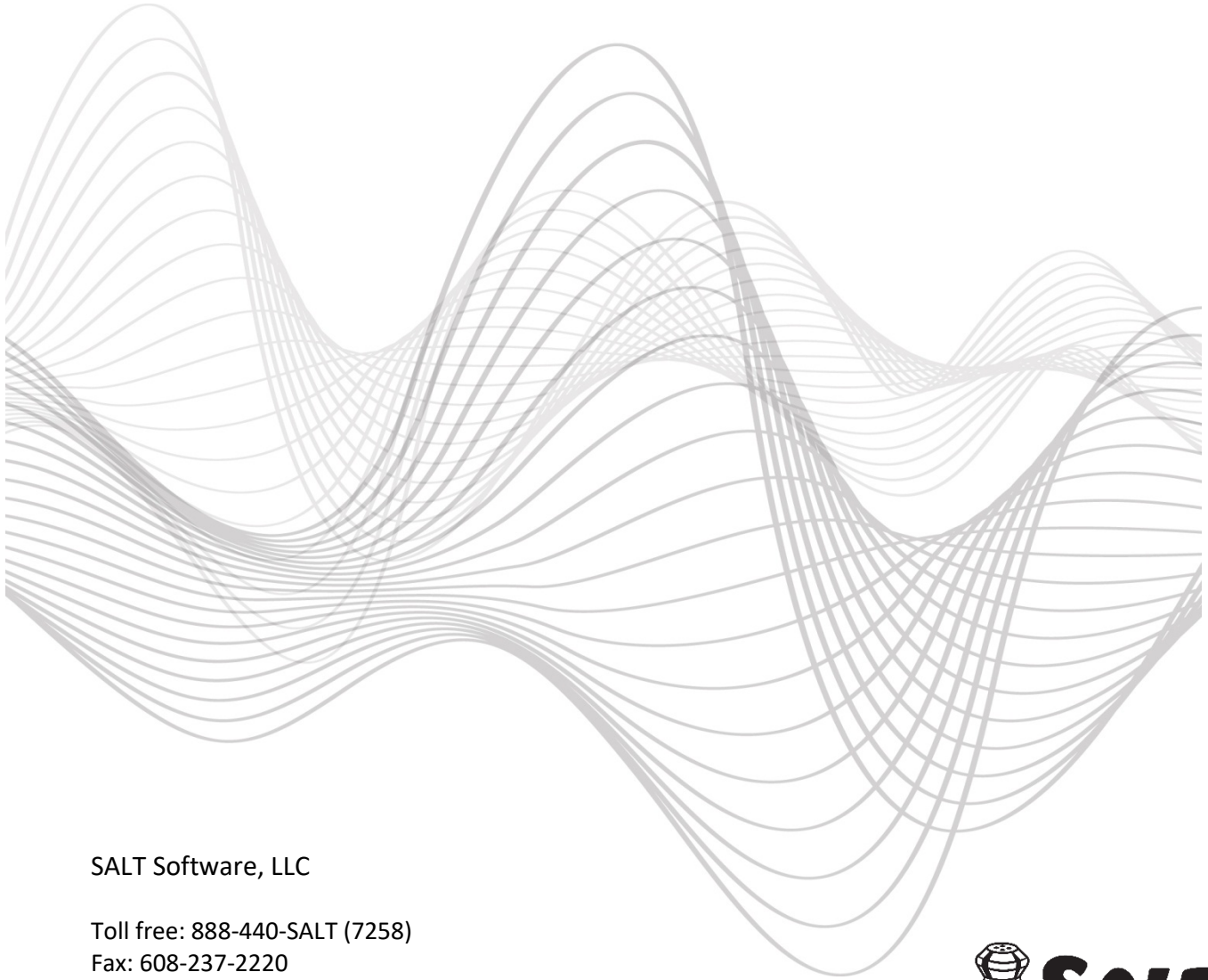# TRANSCRIPTION CONVENTIONS
Adapted for French

SALT Software, LLC

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

Acknowledgments

Table of Contents

A. Using SALT with French Samples

1. **Overview**

   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., *boit|boire, tiens|tenir*.

2. **Character Sets**

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

   - **Windows OS 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. French Conventions – E. Thordardottir
Elin Thordardottir, Ph.D., McGill University

1. INTRODUCTION

The coding system presented here was developed by Elin Thordardottir (Elin Thordardottir, 2005). It has been used and further developed in numerous studies on monolingual and bilingual speakers of French, conducted in the Child Language Development and Disorders Lab (directed by Elin Thordardottir), in the School of Communication Sciences and Disorders at McGill University. Normative data are available for young monolingual speakers of Quebec French (Elin Thordardottir, 2005) and further norms on older children are forthcoming (Elin Thordardottir et al., in preparation). The system was originally developed for Quebec French and available normative data have been collected in Quebec. The coding system itself is equally applicable to European varieties of French. However, differences may be expected notably in lexical use and the comparability of MLU values across French-speaking regions is presently unknown.

This French coding system follows SALT conventions in terms of the elicitation and recording of language samples, communicative context, orthographic transcription, utterance segmentation, assignment of

superfluous language into mazes, transcription of unitelligible syllabes, the coding of missing elements and error utterances. The English and French systems differ in the particular morphemes that are coded. A greater variety of grammatical morphemes is coded in French.

It should be noted that the French MLU cannot be compared directly to the MLU of English-speaking children of the same age. The French MLU is higher than that of English-speaking age mates. This is because a) French has more grammatical morphemes, and b) the French coding system treats productivity differently than the English system. Morphemes are coded in the French system regardless of evidence of productive use (see discussion in Elin Thordardottir, forthcoming).

2. CODING SYSTEM

The French inflectional system is more complex than the English system. This complexity is reflected in the coding. The following grammatical morphemes are coded in this system:

	TENSE /T_	PERSON /P_	MOOD /M_	NUMBER /_PL	GENDER /_G
NOUNS				plural only /NPL	
ADJECTIVES				plural only /APL	masculine or feminine /AG
PRONOUNS				plural only/PPL	masculine or feminine /PG
VERBS	passé composé /T1 imparfait /T2 futur simple /T3 passé simple/T4 plus que parfait /T5 futur proche /T6 passé antérieur /T7 futur antérieur/T8	correct person agreement /P	imperative/M1 subjunctive/M2 conditional /M3 past participle/M4 present participle/M5	past participle when agreement is required /M4PL or /APL	past participle when agreement is required /M4G or /AG

NOUNS: a code is assigned for use of Number (plural). As in the English SALT conventions, use of the singular is not assigned a code. The code used is NPL (Noun Plural). In French, some nouns assume a different form in the plural, but most nouns do not (e.g. *cheval/chevaux*; *chat/chats*). The use of the plural is signaled by the change in the article from *un/une* to *des* or from *le/la* to *les*. The article is not coded separately. The article and the noun are together given the code NPL (e.g. *les chevaux*/NPL)

PRONOUNS: a code is assigned for use of Number (plural only, not the singular). French pronouns must also agree with their referent in Gender (feminine or masculine). Neither gender can be considered a more basic form than the other. Pronouns are always assigned a **Gender** code, whether the gender is feminine or masculine, as long as the correct gender form is used. Pronoun codes start with a P for Pronoun, with, in addition, G for Gender and PL for Plural. Examples: *celuici*/PG; *aucun*/PG; *quelquesunes*/PG/PPL.

Notes:
- As in English SALT coding, the personal pronouns (*je, tu , il, elle, on, nous, vous, ils, elles*) are not assigned any codes but are treated as different lexical items. This includes personal pronouns in a verb complement position (e.g., *me, te, lui, moi, eux, leur*). These receive no code.
- Gender-neutral demonstrative pronouns (*ce, ceci, cela, ça*) receive no code.
- Reflexive pronouns receive no code (e.g. *se* lever).
- The invariable relative pronoun *qui* receives no code, but the inflected relative pronouns *lequel, laquelle, lesquels, lesquelles, àlaquelle, auxquels*, etc. are coded.

- The clitic pronoun (which has received considerable attention in studies on Specific Language Impairment in French) is homophonous with the definite article. It is marked for gender and number as a pronoun. Example: *il **le/PG** fait/P; je **l'/PG** ai/P vue/T1/AG; nous **les/PG/PPL** connaissons/P*
- Two-word pronouns are written as one word (*le mien= lemien, à laquelle=àlaquelle*).
- Many French words are regarded as adjectives when they accompany a noun, and as pronouns when they stand alone (e.g., *chacun, tout, certain*) and are coded accordingly. In cases of doubt, it should be kept in mind that which code is assigned does not affect the MLU. It affects only subsequent analyses of diversity or morphemes used, where it may or may not be considered to be of importance.

ADJECTIVES: as for pronouns, a code is always assigned for correct use of Gender, as adjectives must always agree with their referent in Gender. Number is only coded when the adjective is in the Plural. Adjective codes start with A, with, in addition, G and PL, as appropriate Examples: *un chat **gris/AG**; une **belle/AG** fille; **de beaux/AG/APL** enfants/NPL*.

Notes:
- The gender and plural marking on pronouns and adjectives can be audible or not depending on the lexical item (example: *vert/verte*; *gris/grise*; *rouge/rouge*). The code is assigned regardless of audibility as long as the form used by the speaker is correct given the context (e.g., *un chat **gris/AG**; une fleur **rouge/AG***).
- No plural code is given to adjectives that are inherently plural and have no singular form (e.g. *plusieurs*). Gender is not marked for adjectives whose plural form is invariable for Gender (notably the possessive adjectives *mes, tes, ses*).
- Possessive adjectives (*mon, ma, mes, ton, ta, tes*....) are coded for Gender (e.g. *mon/AG chien*) and Plural (mes/APL chiens/NPL), but not for Person (and not for Gender in the plural form, see above).
- Most adjectives are descriptives (adjectifs qualificatifs). Other types of adjectives include demonstrative adjectives (ce, cette, ces), interrogative adjectives (quel, quelle) and indefinite adjectives (aucun, certain, autre).

VERBS: Codes are assigned separately for Tense, Person, and Mood. A given verb is assigned one or more of these codes, as appropriate. As in English SALT coding, no tense code is assigned to the present tense and no mood code to the indicative mood. Also, as in English SALT, the infinitive is not assigned a code. All tenses other than the present have their own code that starts with T, followed by a number (see Table). Person agreement is coded independently of Tense, and is denoted by the code P. Different codes can be assigned to the different persons by adding a number to the P code (1st person singular= P1 etc) but this is optional and does not affect the MLU count. Person agreement is coded for verbs in all tenses, including the present. Verb moods are assigned an M code followed by a number (see Table). The indicative mood receives no code, but the imperative, subjunctive and conditional moods do. A verb in the present tense that has person agreement will have a /P code only. A verb in another tense of the indicative, e.g. the passé composé will have a /P and a /T code. A verb in the subjunctive present will have a /P and and /M code. A verb in the past tense of the subjunctive will have an M, T and P code (all these examples assume that each of these inflections are produced correctly – if not, only those that are correct are coded; those that are incorrect receive an error code, such as if the correct tense and person are used, but the indicative is used instead of the subjunctive).

Notes:
- Person marked verbs in French are sometimes homophonous with other verb forms, such as the infinitive, the past participle, the bare stem, or another inflected form of the same verb. The rule here is that any morphemes required by the context are coded if the form used is correct, regardless of whether that form is homophonous with other forms of the verb (that is, whether it can be established unambiguously that the speaker truly intended to use the correct form and did not confuse it with another form).

- Compound tenses consist of an auxiliary and a past participle or infinitive form of the lexical verb. The person agreement appears on the auxiliary form, and the tense is signaled by the combination of the auxiliary and other verb form. These compound forms are treated as a whole and are assigned a P for person agreement and and T code for tense. It does not matter for the MLU count whether the P code is attached to the auxiliary or the lexical verb (e.g. *il a*/P *donné*/T1 or *il a donné*/P/T1). However, attaching the P code to the auxiliary is helpful if further analyses of person agreement are to be performed.
- The specific codes for all the French verb tenses are listed in Table 1. Only a few of these are encountered in the samples of preschool children. The present of the subjunctive and conditional moods may be encountered in samples of old preschool and young school-age children. The subjunctive mood also has a passé, an imparfait and a plus-que-parfait. These are coded by T1, T2 and T5 (the T1 code is used because this past of the subjunctive is a compound tense like the passé composé of the indicative). The conditional mood has a past tense (T1) in addition to the present.

PARTICIPLES: The present participle is encountered fairly commonly (e.g. sortir *en courant, un chien ayant faim*). It is coded as M5.

The past participle of verbs frequently functions as an adjective, and in those instances, it is coded as such in this system (e.g. une *fille* **endormie/AG**; *un garçon* **blessé/AG**). In some cases, it is hard to decide whether a past participle should be treated as an adjective, notably in sentences in which it clearly refers to the action of the verb: example: *des touristes* **venus** *de loin; de l'argent* **trouvé** *dans la rue.* In such instances, Gender and Number marking can be coded as /M4G and /M4PL if the coder is interested in making a distinction between adjectives and past participles*.* These instances are rare and whether the form is coded as a past participle or an adjective in no way impacts the MLU count, since the same rules of Gender and Number agreement apply.

In French, the past participle form of compound tense verbs must agree in Gender and Number with its referent under certain circumstances, which are relevant to the use of lexical objects and object clitics (agreement is required: a) always when the auxiliary is the verb *être*, and b) with the verb *avoir* if the direct object is placed before the auxiliary). As for adjectives, the agreement is visible in writing, but is not always audible. However, if the form that is used is correct, the required agreement is always coded. Example: Il avait/P/ **acheté/T5** des pommes et il les/PG/PPL avait/P **mangées/T5/AG/APL.** As with any use of the past participle, the agreement of the past participle can be coded as /AG and /APL or alternatively as/ M4G and /M4PL, depending on whether the coder is interested in keeping track of the distinction between adjectives and past participles, or more interested in keeping the coding system as simple as possible.

Very young children may use verbs with no subject that could be a past participle or an infinitive (e.g. *tombé!*). These are potentially primitive versions of a passé composé, but given that the auxiliary is not present, such verbs are not assigned a grammatical morpheme code.

OMISSION AND COMISSION ERRORS

In SALT analysis, omissions are signaled by * such that omitted morphemes receive a */ code specifying the morpheme that was omitted. In this system, verb inflections are considered omitted if the infinitive is used instead of a conjugated form. If, however, an incorrect conjugated form is used, the * code is not assigned. Instead, an error code using brackets ([ ]) is used. Unlike English, French morphological errors are not overwhelmingly omissions. Therefore, error codes take on more importance than they typically do in English coding.

For nouns, the only possible omission is the use of the singular for the plural (*/NPL). The use of the

plural for the singular is a comission error ([ENN] error noun number).

For adjectives and pronouns, use of the wrong gender is always considered a comission error, since these words cannot have no gender. Incorrect use of the singular and plural is treated the same way as for nouns.

For verbs, omission errors are instances in which the infinitive form is used in the place of a conjugated form. These errors receive the appropriate code preceded by an asterisk (*/P, */T2 etc.). When children inflect a verb incorrectly, an error code in brackets is used instead to indicate the nature of the error.

Common Error Codes: In this system, the tradition is to assign error codes as appropriate that are descriptive of the nature of the error. Depending on errors that are encountered, coders can make up their own codes, as needed. In this system, the following are used:

    E = error
    V = verb
    A = adjective
    P = pronoun
    N = noun or number
    VT = verb tense
    VP = verb person
    W = word
    WO = word order

Error codes that are commonly encountered include:

    [EVT]   incorrect verb tense
    [EVP]   incorrect verb person
    [EAG]   incorrect gender of adjective
    [EPG]   incorrect gender of pronoun
    [ENN]   incorrect number of noun (plural for singular)
    [EW]   incorrect choice of word
    [EWO] incorrect word order

Other codes are formed in an analogous manner as needed.

TABLE 1: BOUND MORPHEME CODES

    /T1   passé composé (present perfect) *j'ai mangé*
    /T2   imparfait (imperfect) *je mangeais*
    /T3   futur simple (future) *je mangerai*
    /T4   passé simple (simple past) *je mangeai*
    /T5   plus-que-parfait (past perfect) *j'avais mangé*
    /T6   futur périphrastique (periphrastic future) *je vais manger*
    /T7   passé antérieur, *j'eus mangé*
    /T8   futur antérieur, *j'aurai mangé*
    /P   verb person marking *je mange, tu manges etc*
    /M1   impératif (imperative mood) *mange!*
    /M2   subjunctive mood) *je ne veux pas que tu **manges,** il faut que tu le **saches***
    /M3   conditionnel (conditional mood) *si j'avais faim, je **mangerais***
    /M4   past participle, *des plats **mangés** de tous; des touristes **venus** de loin*
    /M5   present participle , *en **mangeant***
    /PLN   noun plural, ***les** oiseau**x**, **les** voiture**s***
    /APL   adjective plural, *les **belles** choses, des livres **intéressants***
    /PPL   pronoun plural   ***toutes** les personnes*

/AG   adjective (and past participle) gender agreement *la voiture **grise**, le **gentil** voisin*
/PG   pronoun (and clitic) gender agreement *je **la** connais, **chacun** des hommes*

<u>TABLE OF SPELLING CONVENTIONS</u>
Many French phrases and expressions are written as two or more words, but are felt to be a single entity by speakers of French. Accordingly, and consistent with English SALT rules, such words are consistently spelled as one word and counted as one lexical item. The following table lists many commonly encountered words of this type. This table was compiled over many years of coding large numbers of samples. However, it is not exhaustive. If additional multi word entries are encountered that are analogous to the words in this table and should be spelled as one word based on similar principles, then they should be treated as the words in this table.

à cause	àcause		en train de	entrainde
à côté	àcôté		est-ce	que, estceque
ah ben	ahben		fait que	faitque
à la place	àlaplace		il ya a	ilya, coded ilya/P
à laquelle	àlaquelle (and all its forms)		il y avait	coded ilya/P/T2
à terre	àterre		il y en a	coded ilya/3 en
aujourd'hui	aujourdhui		là bas	làbas
Bob l'éponge	bobléponge		là dedans	làdedans
Burger King	burgerking		le mien	lemien
bye bye	byebye		n'importe qui	nimportequi
celui-là	celuilà (and all its forms)		oh là là	ohlàlà
je sais pas	chépas (when pronounced as chépas)		parce que	parceque
coup de pied	coupdepied		par terre	parterre
coup de poing	coupdepoing		peut être	peutêtre
d'abord	dabord		quleque chose	quelquechose
d'accord	daccord		quelqu'un	quelquun
de même	demême		qu'est-ce que	questceque
en arrière	enarrière		qui est-ce qui	questcequi
en bas	enbas		s'il vous plaît	silvousplaît
en dedans	endedans		toutà coup	toutàcoup
en dessous	endessous		tout de suite	toutdesuite
en dessus	endessus		tout le monde	toutlemonde
en haut	enhaut		un deux trois go	undeuxtroisgo
en l'air	enlair			

<u>Notes</u>:
pour que is written in two words
The pronoun *il* is often pronounced as *i.* This can cause ambiguity, which must be resolved using the context. Examples:

y'a du monde	ilya/P du monde
y'a pris une boîte	il a/P pris/T1 une boîte
y'en a deux	ilya/P en deux {= il y en a deux}
y'en a pris deux	il en a/P pris/T1 deux

3. **REFERENCES**

Elin Thordardottir (forthcoming). Typical development and Primary Language Impairment in monolingual French-speaking children. In J. Patterson and B. Rodriguez (Eds.), *Multilingual*

perspectives on child language disorders. Multilingual Matters.
Elin Thordardottir(2005). Early lexical and syntactic development in Quebec French and English: Implications for cross-linguistic and bilingual assessment. *International Journal of Language and Communication Disorders, 40,*243-278.

c. Identifying the Root Form of Verbs and Other Inflected Words

Procedure
- The root identification convention "|" instructs SALT to consider the word immediately preceding the "│"as the main body word and the word immediately following to be the root word (infinitive verb). It is the root word that is used for calculating measures such as NDW (number of different words) and TTR (type-token ratio). Examples: *j'ai|avoir, nous avons|avoir*. The use of *|avoir* ensures that both forms *ai* and *avons* are considered to belong to the verb *avoir*
- An automatic root identification file is in place in SALT for French verbs. For other inflected words, this information needs to be typed in the sample
- Auxiliary verbs and main verbs are both followed by the root identification "|" convention.

Automating Root Identification of Verbs

The root identification file (RIF), <u>French Verbs</u>, contains almost 264,000 verbs with their corresponding root forms and is selected using the *Setup Menu: Language Settings* option. The *Edit Menu: Identify Roots* command looks up all verbs in the transcript that have not been previously identified with the vertical bar. If the verb is not found in the active RIF, that verb is ignored. If only one choice is found, the verb is automatically identified. If a verb contains more than one root option, the user is presented with a list of choices to select from. You should use this automation feature to identify all the verbs in your transcript.

Use of the | Convention with Other Word Classes

The root of French words may change in different inflected forms, including nouns *(cheval, chevaux),* pronouns *(lequel, laquelle),* and adjectives *(beau, belle; vert, verte).* Words that occur more than once in a sample with different root forms must be coded with | followed by a consistent form of the word to ensure that the lexical count does not treat two inflectional forms of the word as different lexical items. Examples: *un cheval, des chevaux|cheval/NPL*. For consistency, the form following the vertical bar | is the singular form of nouns, the masculine form of adjectives and pronouns (which does not, however, presuppose that the masculine form is more basic, only frequently shorter) and the infinitive form of verbs.

Note: Not using the "|" convention has no effect on MLU (mean length of utterance) or NTW (number of total words), but it does impact NDW, the number of different words. The "|" doesn't need to be used if you aren't interested in these measures and other measures based on identifying the root form.