**Contents**

**1 General Principles**

**2 Turn-taking**

**3 Liaisons**

**4 Length of utterance**

**5 Breaks, non-verbal phenomena and variants**

**6 File Naming**

**7 Short-list of transcription conventions**

**9 Word count**

TRANSCRIPTION PRINCIPLES OF THE ORAL CORPUS INTERFRA WITHIN THE PROJECT:

*FRENCH INTERLANGUAGE - DEVELOPMENT, VARIATION AND INTERACTION*

1 General Principles

This report presents the transcription conventions of spoken French varieties of learners and Native French speakers applied within the project *French interlanguage - development, variation and interaction (INTERFRA)*.

(See www.http://www.fraitaklass.su.se/english/interfra)

In general, the transcriptions were performed as thoroughly as possible. This implies that the hesitations, filled pauses, repetitions, etc. were accounted for. We opted for an orthographic transcription of this variety of spoken French (see Blanche-Benveniste & Jeanjean 1987).

This transcription model aims to render the grammatical structure of texts, the planning of a statement and the structure of the dialogue in this exolingual communication. Lexical and morphological forms of words are presented as clearly as possible. Prosody is not specified but it can be an additional criterion for syntactic or pragmatic analysis. (cf. Hammarberg 1993).

2 Turn-taking

a) Turn-taking is indicated by a new line for each speaker and each response: E = student, I = interviewer.

 (1) I: comment as-tu appris le français ?

 E: d’abord à l’école / pendant six ans . (Interview 1, Marie)

b) During simultaneous speech (between the two interlocutors), the + sign indicates the beginning of the simultaneous speaking and “SIM” indicates the end of the simultaneous contact of speech. A space is left between "+" and the following word. By marking thus, the transcriber calls attention to the simultaneous contact.

 (2) E: ah non / + je sais pa:s .

 I: tu n’as SIM jamais pensé devenir professeur de français ? (Interview 3, Yvonne)

c) When the interviewer provides supportive "feed-back" such as the sound *mm* in the interaction with the subject, it is not counted as a full turn. Such “feed-back” is indicated in parentheses within the utterance of the subject, possibly followed by SIM

 (3) E: j’avais une / une image euh euh très euh stéréotypée des Français et de la France / *(I:mm)* alors / que maintenant je j’ai appris beaucoup de choses de la société par *(I:mm)* SIM qui est // (Interview 5, Marie)

Note that there must be no space between I: / E: and the text in parentheses.

d) These markers are rendered "mm" or "mhm."

e) The feedback may also consist of words such as "yes," "no," or similar responses. They are noted similarly in the turn of the second interlocutor. These expressions can, of course, be a full turn and are then marked as a new turn.

f) It is sometimes difficult to decide to what extent these expressions are separate turns, or if they simply fill a support function of speech and should therefore be included in the text. One way to solve this in ambiguous cases, is to consider how the conversation continues after the marker "mm", if the learner has completed his reply, we write "mm" in the reply of the interviewer. In this case, "mm" is not considered a real support utterance. Similarly, if the function of "mm" is to interrupt a turn, to speak or to change the subject of conversation, then it is counted as the beginning of a new turn and introduces a new turn.

g) Note that only SIM is used when the support markers are simultaneous. The + sign is only used when it comes to turns that overlap (cf. ex. 3).

3. Liaisons

a) The liaisons are not marked when they are obligatory. Here we count the following:

un ami, mon enfance, on est, ils entrent, ils ont, vous avez, dans un, quand on, les autres, le petit enfant, vingt ans, très étonné, en avion.

b) The liaisons that are not included in the examples above are considered non-obligatory and are followed by the sign +. For example:

je suis+ allé, ils sont+ un peu, c’est+ une ville, c’est+ intéressant, c’était+ un lanceur de couteaux, c’est+ à toi, pas+ encore, elle est+ assise, mais+ il y a, il faut+ avoir etc.

The + sign is also used when the liaison is obligatory and an element appears between the two words linked by a liaison.

 (4) très+ euh / excités

The degree to which these liaisons are optional warrants discussion. A review of several studies of spoken French and an analysis of the native French speakers showed that in the cases mentioned under a) the liaison is made without exception. For this reason we have chosen to classify all other liaisons as "non-obligatory." The project's goal is not to study the frequency of liaisons at different levels of styles or registers, so a sort of "lowest common denominator" has been established with respect to liaisons, cf. a).

Therefore we only indicate learners' omitted liaisons when they deviate from typical cases where the liaison is obligatory. In this case we use the hyphen with a space on each side so as not to be confused with hyphens indicating repetition (cf. 5):

 (5) et les - autres = [le otR]

c) Consequently, sounds that are not pronounced are indicated, e.g. final consonant:

 (6) quarante - points (Interview 1, Anders) = /t/ est omis dans *quarante*

 (7) tout le monde - (Interview 1, Anders) = /d/ est omis dans *monde*

Or, when the end of a word is not distinct:

 (8) une semaine -

d) In contrast, the + sign indicates that a silent, final consonant is pronounced:

 (9) elle est assis dans un banc+ (Vidéo 1, Marie)

 (10) et elle met+ une un \*scarf euh // (Vidéo 1, Yvonne)

e) The overuse of a liaison is marked as follows:

 (11) il y a deux hommes dans l’ambulance . ils prend z’un // (Vidéo 1, Anders) = You can hear a the sound [z].

4. Length of utterance

Within the project INTERFRA, we hold that the international measure MLU (Mean Length of Utterances established by Brown (1973), which is used to determine the complexity in the language production of children) is not necessarily relevant as criterion for adult learners. High MLU does not seem to correlate with a greater complexity of the language of acquisition. It seems that long sentences with several subordinates may be a kind of "avoidance strategy". A subordinate clause can, for example, be a paraphrase for a word that the learner does not know (cf. Bartning 1990).

We have chosen to put a full stop in the production of the learner to mark macrosyntagms in our analysis. In this regard we refer to the definition of macrosyntagm by Loman / Jörgensen (1971). A macrosyntagm is the smallest part of a text which is syntactically independent. That is to say, a sentence with its possible subordinate clauses. In other words, "a sequence of segments, internally linked by a network of syntactic relations and externally bound by the fact that such relations do not appear between adjacent segments" (Loman & Jörgensen 1971, p. 9). Therefore, it is firstly a syntactic criteria determining where the period should be. Prosody (pauses and information, e.g.) is also to be considered for the syntactic interpretation.

a) Therefore the punctuation mark of a period to be marked in the learner's production after

1. A macrosyntagm containing at least one finite verb

(12) E: e:t / eh l’enfant qui qui ne pèse pas e:h trop / eh il monte avec les ballons parce que il est il est il n’est pas au- a- assez lourde pour pour les / les tenir au monde . (Video 2, Lena)

2. After a fragmented phrase i.e., a part of a phrase without finite verb, interpreted as an independent clause and which does not constitute a "launch pad" or a reformulation (cf. 13 below).

 (13) I: ah bon .

 E: oui des études de français . (Interview 1, Eva)

From a syntactic point of view, a macrosyntagme (MS) may therefore correspond to a sentence in written language. But other than the sequences constituting the internal syntactic relations of a MS, you also find between two points the sequences reflecting the planning of the utterance: pauses, hesitations, repetions, self-corrections and other phenomena typical of spoken language. In this way, we can understand both the features of sentence planning and those of the resulting sentence (cf. Hammarberg 1993, p. 2).

b) Use of the # sign

Two types of utterance planning can be distinguished. The first type involves the reformulations which are indicated by the # sign. The other type involves self-corrections (cf. c) below). Reformulations may be the product of one’s own initiative or can occur during a speech turn which results in an incomplete MS.

1. Reformulation of one’s own initiative

Here we mean syntactic reformulation of the utterance, with or without breaks "//" and hesitations like "étaient" as seen in example (14):

 (14) E: et on discutait des choses qui étaient # // s’il y a des sujets de matières que je m’intéresse à ça va . (Interview 3 Yvonne).

 (15) E: / puis je # il y a pas beaucoup de: / temps . (Interview 3, Yvonne)

 (16) E: je travaille da:ns # j’ai deux travail alors + s . (Interview 3, Yvonne)

It can sometimes be difficult to determine whether it is a reformulation ("#") or an MS ("."), which appears in the following example:

 (17) E: et on discutait des choses qui étai:ent # // s’il y a des sujets de matières que je m’intéresse à ça va . *mais s’il y en a d’autres* . mais c’était très très bien parce que les animateurs étaient bien e:t . (Interview 3, Yvonne)

The sequence in italics may at first appear as a reformulation. However, it is not a syntactic reformulation of the same utterance, since the continuation indicates a change of subject; the student continues the statement speaking about the teachers of the course instead of the conversation topics. It is rather an ellipse "mais s'il y en a d'autres, ça va pas" or something similar. The sequence contains no breaks or hesitation.

2. Unfinished MS

This type of reformulation appears during a turn which causes the interruption of the first interlocutor's MS:

 (18) E: je suis toute seule là . // *mai:s #*

 I: c’est plus fa- c’est plus agréable de: servir le café à Dramaten ?

 E: oui je pense oui . (Interview 3, Yvonne)

If the turn includes a simultaneous utterance the # sign is not used, since the symbols "+ [...] SIM" already indicate why the first MS is incomplete:

 (19) E: pas tout le temps je pense + à: (RIRE)

 I: pas tout le temps . SIM // les spectateurs sont aimables .

 E: oui c’est ça . (Interview 3, Yvonne)

N.B. If the first interlocutor ends his turn after the intervention of the second interlocutor, so that the MS continues on two turns, the # sign is not used either, because the MS is not interrupted.

 (20) E: c’était sa- samedi soir + (RIRE RIRE)

 I: ah oui ah ben oui SIM / d’accord .

 E: à neuf heures . (Interview 2, Kerstin)

c) Self-correction

This other type of planning statement includes the repetition of identical words or a sequence, slips and self-correction. These are not followed by the # sign because they do not entail syntactic reformulation of the utterance:

 (21) E: toujours *le petit le petit* rentrée le petit *plat plat* de résistance le / fromage et / dessert comme ça . (Interview 2, Anders)

 (22) E: mais dans une cours euh pour *les corre- euh / conversation* c’est très bien *qu’il ne fa- / qu’ils ne fassent* pas trop + quand même .

In cases where fragmented sentences or reformulations make the utterance ambiguous, prosody helps the syntactic analysis:

 (23) I: et ça va se passer quand ?

 E: demain (=intonation descendante). / il arrive tôt demain matin .

 (24) E: demain (= intonation montante) # il arrive tôt demain matin .

The interruption criterion precedes the criterion of the finite verb (cf. a) above). Therefore the end of the MS is not marked after the finite verb in example (25), since the learner restates and continues on the same launch pad:

 (25) je peux pas # je veux pas me prononcer là-dessus.

Note, however, three cases of self-correction that are not considered reformulations:

a. Change of tense

 c’est c’était bon

 il travaille e:h il a travaillé

b. Change of modality

 c’est c’est pas bon

c. Articles, prepositions, pronouns:

 un cours de peinture de la peinture

 cet été avec pour les \*social-démocrates

 il elle il mange

In general, the # sign is only used to indicate a syntactic modification of the utterance.

d) Coordinating conjunctions, especially *et* and *mais* are generally considered as the beginning of a new MS when followed by a finite verb:

 (26) E: je ne lis pas tout / . *mais* je regarde tout . (Interview 1, Marie)

Unlike:

 (27) E: Alice m’a / m’a demandé ça aussi *mais* non (RIRE) . c’est normal pour les Français de faire ça ? (Interview 1, Eva)

 (28) E: un petit peu . / pas beaucoup *mais* un petit peu oui . (Interview 1, Jessica)

If these conjunctions are not followed by a finite verb and if the learner indicates a rising intonation for example that he / she intends to continue the utterance, the conjunction is preceded by a full stop:

 (29) E: je suis toute seule là . // *mai:s #*

 I: c’est plus fa- c’est plus agréable de: servir le café à Dramaten ? (Interview 3, Yvonne)

 (30) E: oui je pense oui . / *mais* pas # (Interview 3, Yvonne)

In contrast, if the learner's falling intonation indicates that the turn ends with a conjunction, the full stop is marked as follows:

 (31) E: mais c’était très très bien parce que les animateurs étaient bien *e:t* . /

 I: mm . / tu as appris beaucoup de choses ? (Interview 3, Yvonne)

 (32) E: j’aime bien la Suède *mais* . // ah je voudrais vivre dans un autre pays . (Interview 2, Jessica)

 (33) E: mais c’est (BRUIT) c’est le seule chose que j’ai remarqué comme ça *mais* .

 I: mhm . / où habitais-tu à Besançon ? (Interview 2, Yvonne)

When there is an ellipse of the finite verb after the conjunction *et*, then the whole sentence is considered a macrosyntagm:

(34) je vais prendre une douche *et* téléphoner à mon père .

When the subject of the second finite verb is omitted, the second phrase is considered to belong to the previous macrosyntagm?

 (35) je fais la vaisselle et *parle* au téléphone en même temps .

Versus:

 (36) E: je fais la vaisselle . et *je parle* au téléphone en même temps .

d) Subordinating conjunctions such as *si* and *que* are not preceded by the punctuation of a period, even if they introduce a new finite verb:

 (37) E: puis elle m’a dit *que* si je veux je peux toujours eh / revenir chez elle . (Interview 5, Marie)

In our material, it is difficult to always consider *parce que* as a subordinating conjunction. In general, however, we do not mark the full stop before *parce que*. The analysis of this connector is a subject of extensive studies within the project (see Hancock 1997). Exceptions to this rule are made in cases where the intonation and pauses indicate that the learner introduces a new MS with *parce que*:

 (38) E: je pense que c’est assez difficile d’éduquer des enfants . / *parce que* on doit pas être trop rigide et pas trop euh / hm st / trop laisser faire . je sais pas . (Interview 2, Yvonne)

 (39) E: mais ça c’était un peu différent parce que c’était pas l’université comme ça . / *parce que* je ne savais pas suffisamment de français pour (I:mm) être à l’université . (Interview 3, Eva)

In cases where we cannot determine the status of *parce que* a period is not marked and common usage is followed which is to analyse this conjunction as a subordinate conjunction of cause. By analogy with *et* and *mais* in the examples (31, 33), *parce que* seems to also finish a MS and is then followed by a full stop. This choice is supported by a falling intonation:

 (40) E: c’est ça que je pensais aux vingt points . mais maintenant (RIRE) / (I:mm) ça n’existe pas *parce que* .

 I: mhm . / et qu’est-ce que tu fais en littérature ? (Interview 4, Yvonne)

e) When the full stop and pause coincide, we put the full stop first:

 (41) mais c’est mon propre appartement . / c’est ça qui est bien (Interview 4, Yvonne)

f) A space is left before and after the period.

g) The punctuation that defines the macrosyntagms are also indicated in the production of the interviewer.

h) The question mark replaces the period after a MS with an interrogative function, e.g. after an interrogative word, interrogative intonation, or and attempt for confirmation (= the *ai-je bien entendu* function). A space is left before and after the question mark. The question mark is always marked in a solicitation, even if it interrupts the MS:

 (42) E: (SOUPIR) je ne sais pas . / euh l- / euh le Suédois peut-être ou euh / les sciences / euh *qu’est-ce-qu’on s’app- ?* / les sciences euh / (RIRE) euh / (Interview 2, Marie)

 (43) E: euh / oui ils sont un peu / *qu’est-ce-qu’on dit ? / patriotiques ? / patriotes ?* / mais / quand même ils / ils / ils voient les fautes . (Interview 2, Marie)

In addition to the examples listed above, note some special cases marking a MS. These are expressions such as: *je trouve, je pense, je crois, je veux dire*, which are grouped under four scenarios:

a. Punctuating \_\_\_\_\_\_\_\_\_\_\_\_. (Concluding comment pronounced with a low tone, or if the statement is part of what has been said).

b. Introducer. \_\_\_\_\_\_\_\_\_\_\_\_\_ (If the utterance is part of the launch pad)

c. Independent MS. \_\_\_\_\_\_\_\_\_\_\_\_\_.

d. Incise \_\_\_\_\_\_\_\_\_\_\_\_\_

a. Punctuating:

(44) E: c’est assez violent *je veux dire* . / on on habitait dans une tour . (Interview 1, Catherine).

 (45) E: je sais pas trop ce qu’il fait (RIRE) (I:non) *je dois dire* .

b. Introducer:

 (46) E: ou du moins le système que je connais celui de Toulouse . / (I:mm) eu:h *je trouve qu*’ici en Suède on on s’intéresse plus eu:h à l’oralité . (Interview 1, Henri)

 (47) E: aussi mai:s *je pense que* peut-être ici comme il fait froid [...]. (Interview 1, Françoise).

d. Incise:

 (48) E: enfin je m’intéressais pas trop à la politique *je crois* / non plus (I:mm mm) SIM à cette époque là . (Interview 1, Catherine).

 (49) E:e:h ils ils ils peuvent en euh de: / en profiter (I:mm) pour SIM pour eu:h pour élever ce leur eu:h *j’sais pas /* intéresse de la littérature ou quelque chose . (Interview 3, Lena).

These remarks also apply to other expressions such as: *alors, enfin, donc, voilà.*

**5. Pauses, nonverbal phenomena and variants**

a) Pauses: / = Pause about two seconds at most. This sign may also designate stunted language.

// = Longer pause about three seconds or more

/// = Longer pause

A space is left if E is slow to respond to a question by I, this is indicated by a pause at the beginning of the turn.

b) The non-verbal phenomena are capitalized in parentheses: (LAUGHTER) (SIGH) (NOISE). If they are simultaneous with the statement, they follow the utterance as in the example below:

 (50) E: euh une vocation un peu tardive oui c’est vrai (EN RIANT) . (Interview 1, Anders)

Hesitation markers are written as they are pronounced, often *euh, eu:h, eh, mm*. Sometimes problems of interpretation may arise because of the similarity between the hesitation mark *eh* and the conjunction *et*. The latter may function as a marker of hesitation, or at least a word that one pauses upon before continuing with the next sentence. In ambiguous cases, the sign X is used for uncertain transcriptions (cf. c) below).

 (51) E: eh  *Xet* elle est placée sur un / sur un mur ou quelque chose . (Vidéo 1, Marie)

Indefinable non-verbal sounds which cannot be classified under any of the categories named above (LAUGHTER, etc. SIGH.) are indicated (NOISE). Exception: Snap of tongue is indicated by *st*, without parentheses or italic.

c) Additional transcription signs:

All text is written in lower case letters except for proper names.

The name of the subject is always written \*\*\* or NN. The same sign can be used for names, for the purpose of anonymity. The subjects’ first names are replaced by pseudonyms without quotation marks.

" " Indicates direct speech

 (52) E: il appelle le garçon . et il dit euh ”voilà j’ai trouvé une mouche dans dans (BRUIT) mon mon verre” . (Bande dessinée 2, Jessica)

"\*" The asterisk is placed before code switching.

 (53) E: il y plus de / des leçons / (I:mm) SIM / des séminaires / \**something*. (Interview 1, Mona)

 (54) \**realia*

 (55) E: ici je pense pas que c’est si euh / eu:h \**utspritt* (RIRE) (Interview 2, Yvonne)

The sign precedes the word in question, without space.

“< > “ In the (G) group, code switching occurs mainly in Swedish. The Swedish material is indicated in brackets.

E.g: <nej>

<va heter de nu på franska ?>

(Note! There is no space between the brackets and the text)

The only exceptions are names, which are not indicated in brackets.

E.g: Stockholms skärgård

Kungsholmens gymnasium

Det stora Blå

Madame Butterfly

"... " Indicates suspension at the end of utterance

A space is left before and after the ellipsis mark. This sign is only used in the production of the interviewer.

 (56) I: et tu voudrais ouvrir un restaurant à Stockholm ou ... ? (Interview 2, Yvonne)

 (57) I: c’étai:t égalité ? chacun faisait sa part de travail ... ?(Interview 2, Yvonne)

":" For example *e:t* = the prolonged sound on which one rests.

 (58) a:vec

 (59) *e:t j’ai:* vingt-cinq ans . (Interview 3, Yvonne)

Capital letters are used for book titles. Each word of the title begins with a capital letter:

 (60) I: tu n’as pas lu *La Goutte D’Or* ? (Interview 5, Eva)

 (61) E: on a fait *L’Avare* de Molière ? (Interview 5, Eva)

Capital letters are also used to indicate emphasis on a word or syllable.

 (62) E: LA grammaire (RIRE) (Interview 3, Yvonne)

X = uncertain interpretation. The sign immediately precedes the word whose interpretation is difficult (cf. example 63). In the case of a sequence, each word of uncertain interpretation must be preceded by X (cf. example 64). If you can not discern the different words, a phonetic interpretation is then made, preceded by X, cf. (65) (66):

 (63) ça je Xcompte pas (Interview 3, Yvonne)

 (64) Xsais Xpas Xc’ Xest (Interview 3, Yvonne)

XXXX = sequence of unidentifiable syllables. The number of X corresponds approximately to the number of syllables.

d) The following abbreviations are written in capital letters and without points:

 (67) E: et aussi à l’IUFM (Interview 2, Anders)

e) Repetitions and hesitations

If the interlocutor repeats one or more of the first syllables of a word, before pronouncing it, i.e. hesitates, the spelling of the word in question is respected. If the syllable may be confused with other words, we mark a hyphen immediately after it, so that this is not confused with the hyphen used for the omission of an obligatory liaison ( cf. b) above):

 (68) a- a- a- après

 (69) en- en- encore

 (70) di- di- dire

 (71) ens- ens- enseignant

f) Variation and ambiguity in the interpretation of a word:

Generally parentheses indicate ambiguity, i.e. several interpretations are possible:

 (72) il(s) prenai(en)t

Parentheses indicate that the utterance can be interpreted both as singular and plural. The context can often guide the interpretation, especially in narrated sequences. We recommend a limited use of parentheses, since they are removed in a subsequent computer processing. On the contrary, the following is transcribed:

(73) ils prend

If the context suggests that the subject has a plural referent and that the verbal form differs phonetically between singular and plural.

Other cases:

E = indicates the variations of the phoneme /e/, i.e. several interpretations are possible

 (74) passE = passé / passait

 (75) jE = j’ai / je

 (76) des / dE

 les /lE

Here, the context helps the interpretation. If it may be plural, *des / les* are used.

 (77) E: il n’y a pas *des* / *des* conversations . (Interview 1, Mona)

But on the contrary:

 (78) E: c’est // presque tien *dE* pédagogique . (Interview 1, Mona)

Rendering phonetic variants is not the primary purpose of the project, except in cases where they may cause a difference of meaning, cf. examples (74-76). If the pronunciation of the learner deviates markedly from that of the current use of the target language, the phenomenon can be marked by capital letters, which, however, should not change the spelling of the word:

 (79) E: c’était un grand / eh / *CHanGE* cha change de niveau . (Interview 1, Anita) = the sound "ch" is voiced and "ge" is voiceless.

 (80) E: pour *Soigner* l’arbre (Vidéo 1, Eva) = voiced “s”

 (81) E: *commEntaires lIngUistiques*  (Interview 1, Mona) Non-nasal vowel = in the first two cases and Swedish pronunciation “gvi” of the “gui” syllable.

g) Sound reduction

Since sound reductions are a very common phenomenon in spoken French and may possibly show some mastery of the language we have marked them in the following way. If *il* or *ils* is pronounced *i*, it is marked *i.* In the case of a silent *e* it is marked as shown in the examples (82-84):

 (82) je n’veux pas .

 (83) je m’demande

 (84) j’sais pas (= pronounced "ch" sais)

This phenomenon can be indicated by writing as if it were an elision before a vowel, cf. *je n'aime pas*.

**6. Naming of files**

During transcription, Word Perfect files are named according to the following principles:

1. Group

L = Learner of the longitudinal study

T = Learner of the cross sectional study

R = Learner of the future teachers’ group (Rennes group)

C = Control subject (Native French speaker)

G = Secondary school students

N = Beginners

D = PhD students

F = FSL juniors

Q = FSL seniors

M = Multi-task group (Near-native speakers)

J = Control group of junior native speakers

S = Control group of senior native speakers

K = Multi-task control group (Near-native speakers)

2. Activity

Int = Interview

BD = Comic strip

Vid = Retelling of film in French

Vis = Retelling of film in Swedish

Ret = Retrospection

The figures that follow these activities indicate the time of registration, e.g. Int1 = first recording of the interview, BD2 = second recording of the comic strip etc.

3. Pseudonyms

The first three letters of the pseudonym are indicated:

E.g.: LInt2Mar = second interview with Marie from the longitudinal study.

       TBD 3 Gun = Retelling of comic strip performed during the third semester by Gunnar, a learner from the cross sectional group.

The names of the PC Beta Files (i.e. the files that have been tagged) are reduced to four signs, e.g. LI2M, TB3G, etc.

The reason for this method is that during further processing it may be necessary to "extend" the name so that the file name indicates what type of analysis was made.

E.g.: ELI3YVA = Third interview with Yvonne from the longitudinal group, including only the production of *E:* (E = student) (for example, verbal forms listed in alphabetical order).

Each transcript file starts with a header which consists of the following:

[8909 LInt3Yvo

TRANSK: MK 9301

KOLL: IB 9310]

1st line: file name followed by the recording date.

2nd line: initials of the transcriber and the date of the transcription.

3rd line: initials of the controller and the date of the control.

**7 A Short-list of transcription conventions**

\*\*\* Family name of the student

/, //, /// Pause: short, medium, long

. Period marks the end of macrosyntagm

+ SIM Marks the beginning and end of utterances that overlap

les - autres Absent liaison

met+, il faut+ avoir Pronounced silent final consonant or non-obligatory liaison

en- encore Marker of a repeated syllable if it can be confused with an autonomous word

\* Realia Mark of a foreign word not belonging to the French vocabulary

# Interruption or restructuring

(LAUGHTER) Non-verbal noise (in capitals in parentheses)

st Snap of tongue

eh euh Hesitation sound

X Incomprehensible syllable or uncertain interpretation

XXX Unidentifiable sequence

... At the end of a suspended macrosyntagm

: Prolongation of vowel.

LA grammaire Capital letters mark accentuated syllable

friGidaire Mark of improper pronunciation

il(s) prenai(en)t Orthographic alternations

passE (= passé/passait) Orthographic alternations

$ End-of-text mark

**9 Counting entities 'words'**

The word count was conducted to evaluate the production (quantitative) of the subjects. It also allows you to perform some statistical calculations (frequency of use etc.). This count was made in the WP program and follows the rules it sets forth. Regarding the definition of entities:

- Each graphical unit between spaces is counted as one entity.

- Compound words with hyphen are counted as one word (e.g.: compte-rendu = 1 word).

- The apostrophe separates two words (e.g.: l'enfant = 2 words).

- The unfinished words or truncated words, are counted separately (ex: ven- ven- vendeur = 3 words).

- Multiword sequences are also counted separately (ex: vendeur de ballons = 3 words).

Hence, counting is not intended to be representative of the actual number of 'words' but is considered a tool for comparing and evaluating the amount of language production.

The count is made from transcription files after removing the parentheses that indicate the sounds of hesitation, filled pauses etc. (e.g.: “mm” “uh”) and the interlocutor designations

“I :” and “E :”.

**Bibliography:**

Bartning, I. 1990. « L’acquisition du français par des apprenants universitaires suédois - quelques aspects ». *Revue Romane* 25 :2, p. 165-180.

Bartning, I. 1997. « L’apprenant dit avancé et son acquisition d’une langue étrangère- tour d’horizon et esquisse de la variété avancée ». *A.I.L.E* 9, Paris.

Blanche-Benveniste, C. & Jeanjean, C. 1987. *Le français parlé: Transcription & Édition*. Paris : Didier.

Brown, R. 1973. *A First Language. The Early Stages.* Cambridge, MA : Harvard University Press.

Hammarberg, B. 1999. Manual of the ASU Corpus. Department of Linguistics, Stockholm University.

Hancock, V. 2000. *Quelques connecteurs et modalisateurs dans le français parlé d’apprenants avancés. Étude comparative entre suédophones et locuteurs natifs.* Doctoral Thesis. Stockholm University.

Loman, B. & Jörgensen, N. 1971. *Manual för analys och beskrivning av makrosyntagmer.* Lund : Studentlitteratur.

<http://www.fraitaklass.su.se/english/interfra>

<http://aile.revues.org/?lang=en>