Chapter 13

F-LARSP: A Computerized Tool for Measuring Morphosyntactic Abilities in French

Christophe Parisse, Christelle Maillart & Jodi Tommerdahl

1. Introduction

The version of LARSP which has been adapted for use for French-speaking children has followed the lead of Bol & Kuiken (1990) for Dutch in accurately linking stages of language development with chronological ages. Their approach used two criteria for determining whether a given structure should be included on the chart, and if so, at which stage: the structure should be used by at least 50% of the population at a particular stage; and the median of the frequency with which a structure is used should have a value of at least 1.0. For the French adaptation (F-LARSP), a large corpus of child language in French was analysed to determine at what stage structures should be placed on the new chart and how many of these structures should be included. Further details of the adaptation can be found in Maillart, Parisse & Tommerdahl. (in press). In the present chapter, we focus on the design and implementation of a computerized system for accurately carrying out F-LARSP much more quickly than is possible with current manual methods.

2. Description of French Morphosyntax

2.1 French basic word order

The canonical word order of French is SVO (subject – verb – object), but exceptions occur. The use of object clitics leads to SOV structure (e.g. *il le mange* 'he eats it'); the relative clause tolerates VS order (e.g. VS – *l'homme qu'aime Marie* 'the man that Marie loves' -- but also OSV *que Marie aime* 'that Marie loves'); and the interrogative structure allows VSO order (e.g. *connais-tu ce garçon?* 'do you know this boy?'). However, in oral French, the canonical SVO order tends to be preserved in interrogative forms using a rising pattern of intonation (e.g. *La fille embrasse le garcon* ? 'The girl kisses the boy?') or the *est-ce qu*e locution (e.g. *Est-ce que la fille embrasse le garçon*? 'Does the girl kiss the boy?'). Null subjects are not permitted (* *neige* 'snows'), except for imperatives; an impersonal subject is required (e.g. i*l neige* 'it snows' or 'it is snowing') (Kail, 1989).

Determiners are located before the noun and have different forms including the demonstrative, indefinite, interrogative, negative, and possessive (e.g. *cette table* 'this table'*, une table* 'a table'*, quelle table?* 'which table?'*, aucune table* 'no table', *ma table* 'my table'). In contrast, adjectives are placed after the noun they modify (e.g. *une table ronde* 'a round table*'*) with the exception of a small set (e.g. *petit* 'small'), which are placed after the determiners and before the noun.

2.2 French morphology

French is a moderately inflected language. Due to historical erosion of endings, its morphology is characterized by significant homophony in the spoken language. Verbal forms with different inflections are pronounced in a similar way (e.g. *je mange, tu manges, il mange, ils mangent* 'I/you/he/they eat'). 94% of the verbal forms are homophonous for several inflections (e.g. *il chante/ ils chantent* 'he/they sing') (Paradis & El Fenne, 1995). In French, the gender/ number of

many nouns/adjectives is made clear only by the determiner. When audible, these inflections are formed by a vocalic (e.g. *cheval / chevaux* 'horse/horses') or consonantal (e.g. *petit / petite* 'little' masculine vs feminine) morphophonological alternation (Dubois, 1965). The vocalic alternations are so infrequent that they tend to be lexicalized. The morphophonological alternation « consonant/Ø » is more frequently used to mark the number of the verb (e.g. *part - partent ; dort – dormant*). This alternation is used in many different morphological mechanisms, such as verbal, nominal, and adjectival morphology, derivation, and liaison (e.g. *un petit enfant* 'a little child' where the *t* of *petit* is pronounced, vs. *un petit garçon* 'a little boy', where it is not).

The article system contains definite and indefinite forms. An article precedes the noun and agrees with it in number and gender. The inflections of number or gender within pronouns and determiner classes are based on a vocalic alternation. Lexical adverbs are often formed from their adjectival form plus the addition of *-ment* (e.g. *calme-ment* 'calmly').

Past events are expressed by different tenses depending on several factors. Finished events occurring in the past are expressed by the 'passé composé' with the auxiliaries *avoir* ('have') or *être* ('be') plus the past participle (e.g. *J'ai mangé puis je suis parti* 'I ate then I left'). Interrupted events occurring in the past are expressed by the 'imparfait' with the stem of the verb plus the endings *-ais, -ais, -ait, -ions, -iez , -aient* (e.g. *je mangeais lorqu'il est arrivé* 'I ate when he arrived'). Upcoming events are expressed by two frequent tenses: 'futur simple' for future events and 'futur périphrastique' for very close future events. The formation of simple future is the same for the three different groups: the stem of the verb plus the endings *-rai, -ras, -ra, -rons,- rez , -ront* (e.g. *je mangerai* 'I will eat'). The 'futur périphrastique' is formed from the addition of the semi-auxiliary *aller* and the infinitive form of the verb (e.g. *je vais manger* 'I'm going to eat'). Complex verb constructions, auxiliaries followed by a past participle, and modals including *aller*

('to go') followed by an infinitive are very productive and frequently used to express aspect and mode.

3. From E-LARSP to F-LARSP

The adaptation of the English language LARSP (E-LARSP) to F-LARSP was made in six steps:

3.1 *Step 1*

The two first steps were carried out in joint meetings of the three authors, who are all bilingual in French and English, with two being native French speakers. Target structures existing in English but not French were identified and omitted from the developing F-LARSP chart. This included structures such as the plural forms of nouns which are present in French writing but not audible in speech, the contraction *n't* from the Word column, the genitive, and the contracted copula.

3.2 *Step 2*

New target structures specific to French were identified as potential items to be added to the chart. Examples include structures containing lexical morphology such as *n&FEM* indicating the feminine form of a noun such as *la patineuse* (the female ice skater), and *dislo-G* and *dislo-D* which represent dislocations (repetitions of noun phrases by a pronoun) to the left and right, as in the following sentences:

<u>Moi</u> <u>je</u> suis contente.

<u>Elle</u> est belle <u>la jeunesse</u>.

Dislocations are frequent in oral language (Blasco-Dulbecco, 1999), even in the speech of young children (Parisse, 2008), and are likely to represent an important step in language development.

3.3 *Step 3*

Software identifying the potential morphosyntactic targets of the F-LARSP was developed by the first author.

Automatic coding of the F-LARSP

F-LARSP automatic processing is itself a three-phase process: (a) part-of-speech tagging, (b) extraction of grammatical forms, and (c) statistical analysis and creation of an editable chart. The final user can intervene between phase (b) and phase (c) to check the accuracy of the extraction of grammatical forms and correct the results from the automatic process. F-LARSP, including charts, statistics and software, is available at <www.modyco.fr/flarsp/>. The present 1.0 version is partly based on the CLAN software from the CHILDES project (MacWhinney, 2000), and on Windows-based separate applications for phase (b) and phase (c). Future versions will be available for Apple Mac OS and other systems, and will integrate all phases into a single piece of software.

*Phase (a). T*ranscriptions of the recordings were grammatically coded for part of speech using CLAN. This software is able to produce a part-of-speech syntactic analysis (Parisse & Le Normand, 2000). It is based on two CLAN commands, MOR and POST. MOR provides all possible parts of speech for a given word out of context, as shown in example (1) below. The whole dictionary used in the MOR command was created by the first author (see <childes.psy.cmu.edu/mortags/>). Examples (1) and (2) below provide an example of an

utterance and all the possible tags that its words can have in French. Elements analyzed by MOR are represented by a main grammatical category followed by a '|', the base of the word (infinitive for a verb, masculine singular for a noun) and, whenever it is necessary, affixes separated by '-' for regular morphology and '&' for irregular morphology (see MacWhinney, 2000). The symbol '^' codes the ambiguity between two categories.

(1) *CHI: je crois qu' ils sont passés là+bas .

     [= I think that they went (have gone) there]

(2) %mor: pro:subj|je&1S v|croire

pro:rel|qu'^pro:int|qu'^prep|que^conj|qu'^adv|que^adv:int|qu' pro:subj|ils&MASC

v:exist|être^v:aux|être n|passé^adj|passé^v|passer-PP adv:place|là+bas .

In (2), two words are ambiguous. One is *qu'* ('that'), which can have a variety of functions, such as referential (e.g. relative pronoun (pro:rel)) and linking (e.g. as a conjunction). The other is *passés,* which can be a noun ('the past'), an adjective ('passed'), or part of a verb construction (a past participle). There are many such ambiguities in French, and a specific CLAN command, POST (created by the first author – see Parisse & Le Normand, 2000), uses the distributional properties of French to propose a best candidate out of the various options from MOR. (It can be used for other languages, and has been in one example successfully adapted to English by Brian MacWhinney. See <childes.psy.cmu.edu/mortags/> for further information about language adaptations available.) An example of the result of POST is given

in (3). In this example, the parser did a fully correct job, but the actual results can vary from 95% to 97% of correct tagging depending on the type of language analysed.

(3) %mor: pro:subj|je v|croire conj|qu' pro:subj|ils&MASC v:exist|être v|passer-PP adv:place|là+bas .

Phase (b). The second tool is applied directly to the output of the first phase. A set of hand-coded rewrite rules was designed and implemented with a tool compatible with the CHAT files format. Rewrite rules are applied first to the results from phase (a) and are recursively introduced until no further rule can be applied. An example of the final result is presented in (4). The results are presented as a text representation of a *n*-ary tree. A sub-tree is represented between square brackets with a similar number, for example [4 @Cop … v|passer-PP 4], line 3 and 4 of example (4). The highest node is [1 … 1]. Elements between square brackets in a node are sub-trees. The other elements belong to the node, such as '@Cop' and 'v|passer-PP' above.

(4) %ctr: [1 @SV [2 PROV [3 @PronP [4 pro:subj|je 4] 3] VB [3 iVB [4 v|croire 4] 3] 2] @Conj [2 conj|qu' 2] @ExpV_SVA [2 PROV [3 @PronP [4 pro:subj|ils&MASC 4] 3] VV [3 @AuxPP [4 @Cop [5 iCop [6 v:exist|être 6] 5] v|passer-PP 4] 3] SmpAdverbial [3 iSmpAdverbial [4 @Locatif [5 adv:place|là+bas 5] 4] 3] 2] . 1]

In example (4), many rewrite rules were used. It is easier to explain how they are applied by following the construction of a specific tree branch. In example (4), the copula *v:exist|être*, corresponding to the word *est* ('is'), is rewritten as 'iCop', which is one of the ways the copula

can be produced in French (it is not a complex copula construction), and this in turn is analysed as a '@Cop', which is the generic name for copula constructions in the rewrite rule system. The '@' which is in front of the construction name means that this construction is to be counted in the final F-LARSP chart, whereas 'iCop' is not, because it is a transient category used only for technical purposes. In turn, the elements '@Cop' and 'v|passer-PP' are grouped together in a '@AuxPP' construction. The '@AuxPP', which will be stored in the F-LARSP chart, is in turn rewritten as a 'VV' which, preceded by a 'PROV' (subject personal pronoun) and a 'SmpAdverbial' (a simple adverbial), makes it a '@ExpV_SVA' (a subject + verb + adverbial construction that includes an expansion of the verb), which is part of the top level description of the utterance, which in our example is '@SV' (subject + verb) @Conj (conjunction) @ExpV_SVA.

The hand-coding approach of this tool has the advantage of offering good control of the identified structure, but is limited in the way that it can handle unusual constructions because handling complex constructions supposes the creation of a very large number of rules. This explains why the tool does not tend to produce incorrect F-LARSP elements, but is not successful in identifying all F-LARSP elements, and especially elements belonging to Stages IV and V. Success is much better for elements from Stage I to III. Nonetheless, after systematic work on all F-LARSP constructions, the tool produces up to 95% of the expected constructions. To facilitate rule design, all computation done on isolated words is performed using a separate set of rules.

Some specific features of French, such as dislocations, were difficult to code by machine. Features of Stage VI and VII from E-LARSP as well as some features of Stage V were

not able to be recognized by the new program due to their variety of possible forms and were therefore not coded. Some of these limitations could be minimized with further improvements in the coding procedure, which will hopefully appear on the F-LARSP website with versions higher than 1.0.

After the process carried out in phase (b), it is possible to edit the results by hand and provide a better final F-LARSP evaluation. The manual process is carried out directly in the intermediate result file from the F-LARSP tool, which contains all necessary information for this, such as the original transcription, the linking with external sound or video if this feature was present in the original file, and the results of the MOR+POST analysis. It is not necessary to correct the tree analysis, but only to suppress, add, or modify information in the tree analysis line that begins with an '@', as the tree information and precise format are not necessary for further processing. The '@' information corresponds to the F-LARSP information relevant to the LARSP analysis and is used during phase (c) (see below).

Phase (c). The third tool is used on the output of the second phase, after manual edition where necessary. All tags generated by phase (b) (automatically or manually edited) that belong to a specific set are extracted and listed. Also, the number of words appropriate to the 'Word' column of the F-LARSP chart is measured. This set of elements is then organized and presented in a format that corresponds to the F-LARSP chart. The format available for the 1.0 version is Excel or OpenOffice worksheets, but PDF presentations could also be computed in the future.

3.4 *Step 4*

The software was evaluated using a large set of language transcriptions to identify and count potential target structures. This corpus of child language was created by Le Normand (1986;

Parisse & Le Normand, 2006). Typically developing participants, all native speakers of French, were recruited from homes and nurseries in Paris and its immediately surrounding areas. This transversal corpus contains 316 recordings of 20 minutes each, from children ranging in age from 2;0 to 4;0. For the needs of this project, additional recordings (Vial, 2010; Dumez, 2010) were carried out of children aged 18 and 21 months.

The corpus was gathered using the following methods. Each child participated in a dyadic interaction with a familiar adult partner (parent or nursery teacher) either in the child's home, nursery, or school. The child and adult were seated at a small table, and the same standardized set of 22 Fisher-Price toys (house, family members, dog, beds, chairs, tables, rocking horse, stroller, cars, staircase) was used with all children. Similar but slightly modified play material was used for the younger children: the material was more recent and from the brand Playmobil®, and some elements were changed to avoid choking hazards. Basic information about the corpus data is shown in the first line of Table 1 below.

3.5 *Step 5*

Accuracy checks compared the software's identification of target structures to those carried out by hand. An error rate of approximately 5% was calculated, mainly consisting of the software's lack of labeling of a structure. When errors detected in Step 5 could be corrected by improving rules used in Step 3, an iterative process (going back to Steps 3, 4, and 5) was used to improve the coverage of F-LARSP features.

3.6 *Step 6*

Statistical analyses were performed on the children's corpus to evaluate the relevance and appropriate stage of all syntactically coded features according to the first standard used by Bol &

Kuiken (i.e. the age that any given structure was used by at least 50% of the population). Results are presented in Table 1. The earliest age at which a structure attains a 50% level or higher is the age that the structure was placed on the F-LARSP chart. Except for a few key items, we have limited the chart to include only structures which the software can identify automatically.

With these results clearly displaying the age-related stage at which particular structures become regular, it is a relatively simple matter to assign examples to the appropriate stage. This motivated the occasional use in our chart of categories produced by at least 30% of the children (see further, Maillart, Parisse & Tommerdahl, in press). As we have developed a program which makes analysing a transcript automatic with regard to finding F-LARSP structures, the time constraint is considerably less problematic, although some clinicians may choose to carry out a hand check for very specific analyses. The automatic F-LARSP processing feature makes the pruning of all less frequent categories from the chart less necessary because infrequent categories do not make the evaluation process more time consuming. For this reason, the inclusion of lesser used structures may be of value in order to perform a more fine-grained analysis of older children or children with unusual production patterns. Of course, categories that are not used at all tend to clutter the evaluation process, so for F-LARSP 1.0 we have included items that never attain 50% but which have been used by a minimum of 30% of children at a certain stage. These categories are identified through the use of italics.

Table 1 contains the results for all F-LARSP structures that were kept in the final chart and the results for all age ranges.

Insert Table 1 about here

4. Discussion

The main objectives of F-LARSP were (1) to provide speech therapists with precise developmental data, which had not previously been available, about the emergence of morphosyntactic structures in French and (2) to provide a complete and rapid assessment tool for analyzing morphosyntactic abilities in oral language. Previously, the coding required for LARSP was very time consuming to carry out by hand, limiting its use in clinical situations. A computerized tool was therefore developed from CLAN transcriptions to reduce the effort involved in carrying out data analyses. This tool allowed us to process a large French-speaking corpus of 329 children aged 18 to 60 months and also to provide precise percentages of children producing structures that corresponded to a particular age level.

The completed F-LARSP chart presented in Table 2 contains, as a reference, the percentage of children producing the structure for the corresponding age level. A usable F-LARSP chart and complete statistics about children's production of LARSP structures will be made available at the web site <www.modyco.fr/flarsp/>.

Insert Table 2 about here (F-LARSP Chart)

One should note that the inclusion of a structure on F-LARSP says little about the number of times the average child uses that structure. A grammatical structure used by 50% or more of the children within an age group was included on the chart regardless of whether the average use was 1 or 30 occurrences. However, a structure greatly used by a few children and not used by most could not attain a place on the chart. Given this, it is possible that the

information provided on this version of the chart is not ideal in its ability to identify individual differences or to suggest important clues about the character of certain language disorders. This problem may be purely theoretical, as work carried out with children with SLI shows that the order of acquisition of grammatical morphemes is similar to typically developing children, with (for example) present tense being acquired before past tense and determiners being acquired before clitic pronouns, regardless of their phonological form (Jakubowicz & Nash, 2001). However, even though the sequence of acquisition is identical, acquisition may still be much later; for example, the omission of articles in children with SLI corresponds to the pattern seen in younger normally developing children (Paradis, Crago, Genese & Rice, 2003). In future it will therefore be beneficial to analyse language samples with language impairments in order to verify if, when compared with typically developing children with the same MLU (mean length of utterance), a similar distribution of categories is found.

Finally, we are aware that adjustments and improvements remain to be made, particularly regarding more complex structures, which have not yet been computerized. At this stage, the software developed is unable to successfully identify elements at Stage V and higher, yet information regarding a child's ability to form complex sentences is vital. For this reason, Stage V of the chart, as well as the dislocations, are being retained in a nearly identical form to that of the original version, and will for the time being need to be carried out by hand. Similarly, Stages VI and VII have been omitted as they cannot yet be identified by the software, and in any case are arguably less informative than Stage V. Further work on developing software in this direction is planned.

As suggested above, the version of F-LARSP described in this chapter is unlikely to be the only one. Future plans for F-LARSP include using it in different settings and with a variety of groups including older children and those with different types of language difficulties, which we expect will provide invaluable feedback from a community of users. Versions using different criteria will then be developed and placed online to provide maximal access and to attempt to meet the needs of different users. These will be accompanied by detailed instructions explaining the construction and use of the chart, how it can be carried out automatically if desired, and how the information provided by F-LARSP can be used in clinical settings.

ACKNOWLEDGMENTS

References

Blasco-Dulbecco, M. (1999). *Les dislocations en français comtemporain*. Paris: Honoré Champion.

Bol, G. W. & Kuiken, F. (1990). Grammatical analysis of developmental language disorders: A study of the morphosyntax of children with specific language disorders, with hearing impairment and with Down's syndrome. *Clinical Linguistics and Phonetics, 4*, 77-86.

Dubois, J. (1965). *La grammaire structurale du français*. Paris : Larousse.

Dumez, M. (2010). *Analyse des conditions de passation et d'application du LARSP en vue de son adaptation francophone*. Mémoire inédit présenté en vue de l'obtention du titre de master en logopédie, Université de Liège.

Jakubowicz, C. & Nash, L. (2001). Functional categories and syntactic operations in (ab)normal language acquisition. *Brain and Language, 77*, 321-339.

Kail, M. (1989). Cue validity, cue cost, and processing types in sentence comprehension in French and Spanish. In B. MacWhinney & E. Bates (eds), *The cross-linguistic study of sentence processing* (pp. 77-117). Cambridge : Cambridge University Press.

Le Normand, M. T. (1986). A developmental exploration of language used to accompany symbolic play in young normal children (2–4-years-old). *Child: Care, Health and Development*, *12,* 121–134.

MacWhinney, B. (2000). The CHILDES project : tools for analyzing talk. Hillsdale, N.J.: Lawrence Erlbaum.

Maillart, C., Parisse, C. & Tommerdahl J. (in press). F-LARSP: An Adaptation of the LARSP Language Profile 1.0 for French. *Clinical Linguistics and Phonetics.*

Paradis, J., Crago, M., Genesee, F & Rice, M. (2003). French-English bilingual children with SLI: how do they compare with their monolingual peers? *Journal of Speech, Language, and Hearing Research, 46*, 113-127.

Paradis, C. & El Fenne, F. (1995). French verbal inflection revisited: constraints, repairs and floating consonants. *Lingua, 95*, 169-204.

Parisse, C. (2008). Left-dislocated subjects: a construction typical of young French-speaking children? In P. Guijarro-Fuentes, P. Larrañaga, & J. Clibbens (eds), *First language acquisition of morphology and syntax: perspectives across languages and learners* (pp. 13-30). Amsterdam: John Benjamins.

Parisse, C. & Le Normand, M. T. (2000). Automatic disambiguation of morphosyntax in spoken language corpora. *Behavior Research Methods, Instruments, & Computers*, *32,* 468-481.

Parisse, C. & Le Normand, M.T. (2006). Une méthode pour évaluer la production du langage spontané chez l'enfant de 2 à 4 ans, *Glossa*, *97*, 10-30.

Vial, C. (2010). *Adaptation française du LARSP. Aspects linguistiques*. Mémoire inédit présenté en vue de l'obtention du titre de master en logopédie, Université de Liège.

Table 1 – Percentages of children producing LARSP target structures

New stage	Old stage	Section type	Age	1;6	1;9	2;0	2;3	2;6	2;9	3;0	3;3	3;6	3;9	4;0
			No. of children	**10**	**13**	**41**	**31**	**37**	**36**	**40**	**34**	**34**	**33**	**32**
			Mean utterances	87	117	73	83	94	110	113	130	117	114	109
			Mean words	114	181	118	185	236	364	397	457	464	475	453
			Category \MLU	1.23	1.38	1.50	2.13	2.44	3.22	3.41	3.51	3.72	3.98	4.01
		Quest	Questions	40	85	73	74	76	89	93	97	94	91	88
	I	Minor	Locative	100	85	83	97	97	100	100	100	97	97	97
	I	Minor	Vocative	100	100	95	100	100	100	98	100	97	94	94
	II	Phrase	Determiner + Noun	40	62	63	100	95	100	100	100	100	100	100
	III	Phrase	Pronoun Other	70	69	73	90	97	100	98	100	97	100	100
	III	Phrase	Pronoun Personal	50	69	73	90	97	100	98	100	97	100	100
		Word	Auxiliaries	20	54	59	81	84	97	93	100	97	97	100
		Word	Past Participle	20	62	66	81	89	97	95	94	97	94	97
I		Word	Present Tense	60	69	78	100	97	97	98	100	100	97	100
	I	Major	Others	0	0	76	90	73	58	50	38	44	33	47
	I	Major	Noun	0	0	90	97	89	75	63	56	65	39	66
	I	Major	Verb	0	0	73	77	84	78	73	79	71	52	59
	I	Minor	Others	60	85	93	94	95	92	98	97	91	91	94
	II	Exp. II	Adjunct	0	31	37	55	70	89	93	100	91	88	97
	II	Exp. II	Object	20	15	41	58	76	89	93	85	85	79	88
II	II	Exp. II	Verb	0	23	49	71	86	94	88	100	94	100	97

	II	Clause	Adjunct + X	20	46	71	81	95	100	100	100	97	100	100
	II	Clause	Subject + Verb	20	46	63	90	95	100	98	100	97	100	100
	II	Clause	Verb + Complement	10	38	76	71	92	97	95	100	97	100	100
	II	Clause	Verb + Object	10	15	49	68	78	92	93	91	74	82	81
	II	Phrase	Preposition + Noun	0	15	39	55	68	97	90	91	94	97	100
	III	Phrase	Auxiliary + Past Participle	0	38	56	71	70	94	90	88	97	94	97
	III	Phrase	Copula	40	54	78	77	92	92	95	91	94	100	97
	III	Phrase	Modal + Infinitive	0	15	37	52	78	94	95	100	94	100	100
	IV	Phrase	Verbal Negation	0	38	56	74	81	89	93	97	94	97	88
	IV	Phrase	Other Negation	0	23	32	58	70	81	78	82	82	73	88
		Word	Infinitive	20	31	61	87	97	97	98	100	97	100	100
		Word	Modal	0	38	51	68	84	94	100	100	94	100	100
	II	Exp. II	Complement	0	8	32	35	73	97	93	97	94	100	97
	II	Exp. II	Subject	10	8	24	39	62	83	78	91	88	85	94
	II	Phrase	Subject Object	10	15	15	29	46	61	55	62	74	58	66
	II	Phrase	Noun + Adjective	10	31	34	45	57	61	58	53	53	52	53
	II	Phrase	Others	10	8	29	45	54	64	68	65	76	85	69
	III	Exp. III	Adjunct	0	0	5	29	54	86	78	94	88	97	88
	III	Exp. III	Complement	0	0	5	13	30	53	50	56	62	61	69
	III	Exp. III	Object	0	8	10	45	73	83	88	97	94	100	94
	III	Exp. III	Verb	0	8	17	32	62	92	83	97	82	97	94
III	III	Clause	Subject + Verb + Adjunct	20	54	49	77	89	94	95	97	94	97	97

	III	Clause	Subject + Verb + Object	0	8	17	48	65	81	90	97	94	100	94
	III	Clause	Verb + Complement + Adjunct	0	0	2	10	19	53	45	47	47	48	47
	III	Phrase	Determiner + Adjective + Noun	10	8	15	48	62	83	80	79	76	88	84
	III	Phrase	Preposition + Determiner + Noun	0	8	20	48	81	94	93	100	97	100	100
	IV	Phrase	Conjunction + X	10	38	24	26	43	61	75	85	82	88	75
	IV	Phrase	Pronoun + Pronoun	0	0	5	23	32	69	68	79	82	85	81
	V	Phrase	Conjunction	20	46	37	42	68	83	90	94	91	97	94
	V	Phrase	Relative Pronoun	10	23	5	19	35	56	68	68	53	70	69
	III	Clause	Subject + Verb + Complement	0	0	15	29	30	39	48	53	44	64	59
IV	IV	Clause	Subject + Verb + Object + Adjunct	0	0	0	6	16	42	63	50	53	55	69
V	V	Phrase	Adjunct + Adjunct	0	0	7	6	11	42	30	44	68	55	47

Additional part (30% threshold)

I	II	Phrase	Noun + Noun	*30*	23	*39*	*39*	*38*	28	15	9	24	*33*	19
	III	Exp.III	Subject	0	0	5	10	11	*42*	28	12	29	*33*	*34*
III	III	Clause	Verb + Object + Adjunct	0	8	5	26	24	*39*	*30*	*41*	24	18	*31*

	IV	Phrase	Preposition + Determiner + Adjective + Noun	0	0	0	3	14	*31*	*35*	*35*	*35*	*39*	*31*
	II	Clause	Verb + Verb	0	0	0	0	8	6	10	*32*	24	12	19
IV	IV	Phrase	X+ Conjunction + X	0	0	0	0	5	19	*30*	12	*32*	18	28
	III	Phrase	Others (3 elts)	0	0	2	6	14	17	15	12	24	21	*38*
	IV	Clause	Subject + Verb + Object + Complement	0	0	2	6	11	19	15	21	24	*30*	*31*
V	IV	Phrase	Noun Phrase + Preposition + Noun Phrase	0	0	0	3	14	25	28	18	*38*	*39*	*31*

Table 2 - F-LARSP Chart

F-LARSP adapted from original by Maillart, Parisse and Tommerdahl (in press)				
	Clause	**Phrase**	**Word**	**Other**
Stage 1 **(1;6-2;0)**		**DN ProO ProP** 62% 70% 69% **NN** 30%	**Aux PP** **Present** 54% 62% 69%	**Q Loc Voc** 85% 100% 100%
Stage 2 **(2;0-2;6)**	**AX SV VC VO SVA** 81% 90% 76% 68% 77%	**AuxPP ModInf PrN** **Cop** 71% 52% 55% 78% **NegV Neg O** 74% 58%	**Inf Mod** 87% 68%	**Exp: A O V** 55% 58% 71%
Stage 3 **(2;6-3;0)**	**SVO VCA VOA** 81% 53% 39% **Dislo-G Dislo-D** 67,7% 83%	**SO NA Others** **DAdjN cX** 61% 61% 64% 83% 61% **PrDN ProPro Conj** **RelPro** 94% 69% 83% 56% **PrDAN** 31%		<u>**Exp: S C**</u> <u>83% 53%</u>
Stage 4 **(3;0-3;6)**	**SVC SVOA VV** 53% 63% 32%	**xCx** 30%		
Stage 5 **(3;6-4;0)**	**SVOC** 31% Coord 1 1+ Subord 1 1+ S O C A	**Adjunct + adjunct** 55% **Others (3elts)** 38% **NP-Pr-NP** 38%		