



Modelling lexical phrases acquisition in L2

Thierry Chanier, Colette Colmerauer, Christophe Fouqueré, Anne Abeillé,
Francoise Picard, Michael Zock

► To cite this version:

Thierry Chanier, Colette Colmerauer, Christophe Fouqueré, Anne Abeillé, Francoise Picard, et al..
Modelling lexical phrases acquisition in L2. Second Language Acquisition Research: The state of the
art, 1992, Utrecht, Netherlands. edutice-00000632

HAL Id: edutice-00000632

<https://edutice.hal.science/edutice-00000632>

Submitted on 29 Sep 2004

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

MODELLING LEXICAL PHRASES ACQUISITION IN L2*

CHANIER Thierry (1), COLMERAUER Colette (2), FOUQUERÉ Christophe (3),
ABEILLÉ Anne (4), PICARD Françoise (5), ZOCK Michael (6)

(1) Département de Linguistique, Université Clermont 2, France.

(2) Laboratoire de Langues, Université Aix-Marseille 2, France.

(3) LIPN-URA 1507, Institut Galilée, Université Paris 13, France.

Email: cf@lipn.univ-paris13.fr

(4) Département de Sciences du Langage, Université Paris 8, France.

(5) CNRS-GRTC, Marseille, France.

(6) CNRS-LIMSI, Orsay, France.

1. INTRODUCTION

The acquisition of lexical competency is considered as one of the major problems in foreign language learning (L2): in order to be able to use properly a given word, the learner must have previously assimilated not only their morphologic and syntactic properties, but also their semantic and pragmatic features. Furthermore, this process of knowledge acquisition is incremental and nonmonotonic, that is, rule formation is based on incomplete data, hence, rules may have to be revised completely in the light of new evidence (data).

The problem we have chosen to focus on within this project is that of *lexical phrase* or *semi-frozen phrase* (SFP).¹ Theoreticians and practitioners of second language acquisition have stressed their importance: they are frequently used by natives in their mother tongue, their acquisition is considered to pose problem in second language learning, and these phrases occur far more often than simple lexical items.²

A SFP is a phrase in which certain parts cannot be altered (these parts being subject to restricted syntactical variations) without affecting the original meaning or function of the expression. Take for example an expression like “abandonner le navire” (abandon the ship). In this case one can substitute the word abandonner by quitter but not navire by bateau.

* This project has been supported by the Parisian joint research board in Cognitive Sciences (GDR 957 "Sciences Cognitives de Paris") of the French national research council (CNRS). M Pengelly of Lancaster University, UK, helped with the translation.

¹ *Semi-frozen phrase* is a translation of the term *expression semi-figée* introduced by the French linguist M. Gross [GROS 82].

² As an illustration, research undertaken at the LADL laboratory [GROS 82] has shown their high proportion within French language (20000 verbal frozen phrases versus 8000 or 12000 free verbs ; 6000 adverbial phrases versus 2000 adverbs ; 300 000 or 400 000 compound nouns versus 80 000 simple nouns).

At first glance SFPs may be divided into two sets: phrases with a *referential meaning*, like idioms, clichés and proverbs [DANL 88; GROS 82; REY 91; COWI 83], and phrases with a *discourse function*, that is, phrases which can only be pragmatically interpreted, e.g. “How do you do?” [MANS 83 ; NATTI 88].³

L2 didacticians provide strong arguments in favour of teaching these phrases [NATTI 88] : the user may operate on larger chunks rather than atomic units (words), hence this strategy alleviates the burden of memory and processing. Furthermore, it allows the learner to focus her attention on conceptual (discourse structure, coherence) and pragmatic aspects of the dialogue (social aspects of the interaction, adequacy of the means chosen). A proficient learner may thus avoid the violation of certain lexical restrictions, errors of register, and so forth when producing a discourse.

Given the fact that the linguistic functions of these two sets of SFPs are fundamentally different, they should be introduced at different stages during the learning process. It would be very difficult to study both of them concurrently, that is why we have decided to concentrate only on one of them, namely, phrases with referential meanings.

An idiom is defined as an expression, whose meaning of its constituent elements does not appear in the global meaning, e.g. *casser sa pipe* (kick the bucket). The non-compositionality of an idiom is generally considered as the discriminating factor of SFPs. The processing of idioms by adults and their acquisition by children have been studied by psychologists. But their studies have been limited so far to L1. Section 2 provides an overview of the results concerning the classification of idioms.

With regards to L2 acquisition of SFPs with referential meaning, we believe that idioms should not be distinguished from the rest, such as clichés: *être gai comme un pinson* / be as gay as a lark. Their learning is often introduced at an advanced level. Mastering such idioms is generally considered to be difficult in production as well as in comprehension, because (a) their meaning cannot be deduced from their components; (b) their syntactic variations are severely constrained (**casser sa propre pipe*⁴); (c) their figurative meanings may introduce nuances obvious for natives, but far from evident for the second language learner (see, for example, the use of *casser sa pipe*/kick the bucket vs. to die). For all these reasons it seems necessary to distinguish SFP acquisition from the acquisition of single words and collocations. For example: *il fait enregistrer ses bagages* / he checks in his luggage. The latter can be semantically decomposed and have no restrictions with regards to syntactic variations. It is perfectly possible to produce variations like:

His luggage was checked in.
He checked his luggage in late.

³ Discourse oriented phrases are indispensable to open and maintain a communication between two speakers (like formulae of presentation, apologise, ...)

⁴ The insertion of any adjective (*propre* here) makes the phrase lose its idiomatic meaning (which we note with a star). Litteral meaning may then be the only one accessible, like to break one's own pipe here.

This paper starts by presenting psycholinguistic evidence concerning the processing and acquisition of SFPs in L1. It provides then a survey of the work done by linguists and a contrastive analysis of SFPs in two languages. Finally, it describes our project from a linguistic, psychological and computational point of view.

The focal points in our study are the following :

- to compare the views psycholinguists and computational linguists have concerning the processes of lexical access and lexical choice;
- to show the similarities holding between the structure of the lexicon in L1 and in L2;
- to offer a pedagogically realistic approach in vocabulary teaching based on these results.

2. ACQUISITION AND PROCESSING OF SFPs

Different researchers have offered different hypotheses concerning the relationships between SFPs and their constituents, and concerning the organisation and access of SFPs within the mental lexicon. The results of this research show that different SFPs are learnt at different moments. These data apply both for learning SFPs in a natural setting or in an institutional environment.

Processing

Early experiments on the comprehension of SFPs were based on the assumption that SFPs are *non-decomposable* and that idioms can be processed either literally or figuratively. It was claimed that the figurative meaning of a SFP is directly represented in the mental lexicon. Each SFP has only one lexical entry, in a similar way as simple words. Authors disagreed on the starting point and the duration of the literal or idiomatic processing.

Lexical access of an SFP is based on its meaning. It occurs either (a) directly before literal processing [GIBB 80]; (b) at the same time as literal meaning is processed [SWIN 79]; or (c) after the rejection of a literal interpretation – this may be the case when contextual factors exclude it as irrelevant [BOBR 73].

These models, like the earlier ones concerning the mental lexicon, are now considered as oversimplified. They cannot explain certain results obtained in more recent experiments. Important differences in the processing of SFPs have been noted. SFPs do not form a unique class of linguistic items. Researchers studied parameters like semantic analyzability [CACC 88; GIBB 89b], familiarity, frequency [SCHW 86; POPI 88], syntactic and lexical flexibility [GIBB 89a].

The assumption that an SFP is semantically non-decomposable (assumption which was used among other things to define SFPs) needs to be reconsidered. Experiments have shown that in order to understand the meaning of a SFP people tend to analyse the expression syntactically and semantically.

The main elements of the phrase are activated during comprehension. The literal and/or figurative meanings of these elements give access to the global meaning even when the phrase

is transformed. It should also be noted, that the role played by the different components is unequal. The stronger the intrinsic meaning of the constituents, the easier will be the understanding of the SFP. The native speakers' intuitions with regards to the potential analyzability of a SFP correlates with its syntactic flexibility, its ease of comprehension and its semantic productivity [GIBB 89a; GIBB 89b].

Cacciari and her colleagues classify idioms in the following way⁵ [CACC 91]:

- type N, non-analyzable phrases: syntactic or semantic analysis of the components is difficult for one of two reasons: (a) the surface structure does not obey the rules of syntax, or (b) the expression contains words which do not exist outside of the SFP. By and large illustrate the former (syntactically odd), spic and span, prendre la poudre d'escampette, apurer un compte, en avoir marre de, are examples of the latter (lexically ill formed).
- type AO, analyzable-opaque: the relationship between the words of the phrase and the meaning of the whole is unclear. However, every word may contribute to the meaning and use of the phrase: kick the bucket / casser sa pipe.
- type AT, analyzable-transparent: the relationship between the words of the phrase and the meaning of the whole is more transparent. This correspondence is based on the fact that each element has often both a literal and figurative (mainly metaphoric) meaning: break the ice/ briser la glace, pop the question, spill the beans, rouler sur l'or.
- type M, quasi-metaphorical: there is nearly a perfect match between the literal and the metaphorical interpretation: abandon the ship / abandonner le navire.

The research mentioned focused on the relationships between the canonical form of a SFP (and its meaning) and the various forms it could take in discourse, or on the methods of analyzing its components. Recently, possible relationships of SFPs belonging to the same semantic field have been studied. For example, Gibbs [GIBB 90] has shown that native speakers use distinctively two semantically similar expressions like blow your stack and bite his head off depending on the appropriateness of the metaphor for the discourse (respectively, "anger is pressurized heat" and "anger is animal behaviour").

We adopt in this study a sort of reverse approach, that is, we start from the global meaning of the SFP and end in its verbalization. The aim is to show to what extent the meaning of a SFP is determined by conceptual metaphors [GIBB 90; CACC 92].

Acquisition in L1

Recent work on L1 acquisition of idioms casts considerable doubt on the predominant view that children are unable to understand idioms before having reached the stage of figurative competence, that is, before the age of 9 or 10 [CACC 89, 92; GIBB 87]. These experiments

⁵ There exist other classifications (such as [GIBB 89a et b]). Cacciari's one is presented here because it has been used for research on acquisition.

also show that exposure, though being very important for production, is not the main factor of idiom acquisition. Furthermore, they show that idiom acquisition does not rely heavily on rote learning.

Levorato and Cacciari identified three stages for acquiring these expressions [LEVO 92]. At stage one (6-7 years), children are able to comprehend the idiomatic meaning if the context provides strong cues. When the child feels that the initial literal interpretation is irrelevant she will replace it with a figurative interpretation, the latter being built by using contextual linguistic knowledge. At the intermediate stage, when the child has enriched her linguistic competency (by then she is able to handle irony, speech acts, conventional metaphors, ...), she may directly access the figurative meaning. It should be noted, that the acquisition of idiomatic expressions requires the comprehension/reconstruction of their figurative meanings when this is possible. The last step is characterized by the fact that the child has full control of the idioms, that is, she starts to use them creatively.⁶ According to the type of idiom (opaque, transparent, quasi-metaphorical) different strategies can be observed.

Other experiments focused more on strategies used by children and adults in L1 in order to understand the meaning of unknown idioms. According to Cacciari [CACC 92] these strategies vary with the type of idiom: "the interpretative strategies children thought as more appropriate reflect the perception of the semantic characteristics and cognitive complexities of idioms". Semantic transfer is used first for quasi metaphorical expressions (*muet comme une carpe* / as dumb as a fish). Transparent idioms require to perform mentally the action described (*chercher une aiguille dans une botte de foin* / to look for a needle in a hay-stack). General strategies like to ask adults, to perform the action, or to find examples, are preferred in the case of opaque expressions (*être au septième ciel* / to be in the seventh heaven). In this latter case, a lot of miscomprehensions occur, which can be viewed as a lack of any strategy.

3. LEXICAL STUDIES IN L1 AND L2

We focus here on lexemes as they are found in a dictionary. Our examples are based on groupings performed by linguists. They range from simple noun phrases or verb phrases to sentences with an idiomatic interpretation. We do not consider proverbs and sayings. In order to gain a better understanding concerning the related semantic phenomena within a language (such as focalisation, irony, register,...) and the syntactic-semantic correspondences between two languages (in our case English and French), we have decided to consider only examples pertaining to a specific semantic field: deception.

N_0 prendre N_1 dans ses filets⁷

N_0 avoir N_1 dans les grandes largeurs / N_0 to pull a fast one on N_1

N_0 doré la pilule à N_1 / N_0 to sugar the pill for N_1

Since meaning has been factored out (it is invariant), we can now show, by varying linguistic and pragmatic factors, how these parameters influence lexical choice.

⁶ Since it involves a much more difficult task, the production competence of idioms appears after the comprehension one.

⁷ N_0 is the deceiver and N_1 the deceived.

Lexical studies in L1

We start with some general, language independent remarks concerning the syntactic structures of expressions.

With the exception of non-analyzable phrases, SFPs have standard syntactic structures, covering simple sentence forms. Gross [GROS 82], after an extensive study of these expressions in French, concludes that frozen noun phrases occur much more often in complement position than in the position of subjects; and that the number of frozen complements is restricted to two, even though syntax may allow for more. The expressions of our field of study (deception) have the same characteristics. Besides this kind of structural information we have to show how context codetermines the meaning and the choice of a specific lexical expression. Hence we have to take into account factors like:

- focalisation: N_0 avoir N_1 / N_0 to have N_1 on versus N_1 se faire avoir / N_1 to be done.
- shades of meaning (nuances of language)
- preservation of the images induced by context.

For instance, the following expressions in French denote different intensities of deception:

1. faire une farce (or variations: faire une bonne/belle/petite/grosse ...)
2. faire une mauvaise farce
3. faire un coup tordu
4. faire un coup fourré
5. faire un coup de Jarnac
6. faire le coup du père François

While the first two expressions convey the notion of jokes, examples three and four express the general notion of deception. The last two expressions add the notion of intensity to the notion of deception. Instead of translating un coup tordu by a bad trick, and un coup de Jarnac by a stab in the back, we investigate how the second language learner marks the degree of deception in the target language.

It is also very important to learn the contexts in which two expressions having approximately the same meaning can be used. The image expressed by a lexical phrase may entail many implications concerning the context (see stylistic classification below). For instance, how does a learner (L1 or L2) become aware of the difference between the following two phrases ?

7. il a trouvé la pilule amère versus la pilule était amère à avaler
or simply la pilule était amère or il a avalé la pilule
8. il a avalé des couleuvres or on lui a fait avaler des couleuvres,
il a eu beaucoup de couleuvres à avaler

Obviously, she soon understands that *swallowing bitter pills* or *snakes* is a rather unpleasant experience that you do only when being forced. Furthermore, she notices that pills are smaller than snakes and easier to swallow. Finally, she learns that a pill is swallowed only once but snakes are swallowed one by one. Note, that pill is always singular, whereas snakes are always plural in the idiom. This is information that has to be highlighted so that the learner

understand when to use which form. For instance, it is possible to help in discovering these constraints by presenting the following examples:

9. Quand il est rentré de vacances il a trouvé son bureau sens dessus-dessous et son ordinateur volé!

When he came back from holiday, his office was all in a mess and his computer stolen !

... la pilule a été amère à avaler !

... a bitter pill to swallow

...* il a avalé des couleuvres !

... he had to swallow snakes

10. Quand il s'est marié toute sa belle famille lui a reproché ses origines paysannes et le lui a fait sentir

When he got married all his family in law resented his countryside origins and let him know

...* la pilule a été amère à avaler !

... a bitter pill to swallow

... il a avalé des couleuvres !

... he had to swallow snakes

Furthermore, a person knowing the meanings of farce / trick, pilule amère / bitter pill and couleuvres / snakes can easily understand the examples (1,2,9,10), but not (3,4,5,6) which are more "frozen" idiomatic expressions.

We agree with Gibbs when he states: "The more analysable an idiom (i.e. the more speakers are aware of these phrases as having separate meaningful units), the more likely that the expression is syntactically productive" [GIBB 89a:102]. And we could paraphrase this idea by "the more learners are aware of these phrases as having separate meaningful units, the more easily they will understand and learn them" provided that there are semantic markers (intensity, frequency, actors, ...).

Without going too deeply into rhetorics [FONT 77], we would like to remind the reader of the definitions of a few figures of speech. A *trope* is a figure of speech that figuratively enhances the explanation and the expression of an idea. There are various tropes, metaphors being the most well known. A *metaphor* is an image that signals a resemblance between the object described and the idea expressed by the metaphor (he is a wolf = this man is bold like a wolf). *Metonymy* is a figure of speech where a symbol stands for the whole (the crown for the monarch), whereas in a *synecdoche* a part stands for the whole or vice-versa (50 heads of cattle for 50 cows). The figures of speech can also be combined.

They exist in any language.

For example, in French un renard / fox is a metaphor, une fine épice / spice is a metonymy and an expression like pas folle la guêpe is a litotes.⁸ The fact that these tropes are linguistic universals is very useful for L2 acquisition.

⁸ A litotes is an understatement. pas folle la guêpe means here: very clever !)

Comparison of phrases in L1 and L2

By comparing languages like Italian, English and French one notices that there are many correspondences between idiomatic expressions, be it at the lexico-syntactic or metaphorical level [D'ELI 90 ; FREC 85]. This comes as no surprise, since these countries are also culturally and linguistically relatively close.

In a comparative study of Italian and French, Conenna has shown [CONN 84] that out of 2000 Italian expressions having the structure V - frozen direct object complement, more than fourty percen could be translated (nearly) literally into French. Freckleton [FREC 85] studied numerous idioms, including the English verbs take, play and hit: nearly eighty percent of the 238 expressions built with the verb take have an idiomatic correspondence in French, 20% have a quasi word-by-word translation.

With regards to play and to hit the corresponding percentages are respectively 77%,15% out of 60 expressions (37%, 3% out of 38 expressions).

Approximately the following correspondences can be given for two languages:

1) Word-for-word correspondences (neglecting marginal differences such as presence of determiners, possessives, matching of number, i.e. singular vs. plural)

14. N₀ cache ses cartes, son jeu / N₀ hides one's cards

15. N₁ tombe dans le piège / N₁ falls into the trap

Other well known examples falling out of our restricted domain are:

break the ice / rompere il ghiaccio / rompre la glace

take the bull by the horns / prendere il toro per le corna / prendre le taureau par les cornes

2) Partial lexical difference, similar tropes

In French jouer au plus fin means to play the smartest wit, wit remaining unexpressed. Although, to have a battle of wits is an expression in English.

16. N₀ joue au plus fin avec N₁ / N₀ has a battle of wits with N₁

17. N₀ traite N₁ à la légère / N₀ plays fast and loose with N₁

Or :

be as gay as a lark / être gai comme un pinson

kill time / passare il tempo / tuer le temps

3) Structural correspondences, different tropes

The same idea is expressed by two "tropes" with the same topic N₀:

18. N₀ joue/ mise sur les deux tableaux

19. N₀ runs with the hare and chases with the fox

Or :

hit the road / prendre le large

This case seems to pose problems for a second language learner, yet an adult will quickly understand, provided that all the vocabulary is explained (hare, fox, chase). However, a

problem may remain in cases with opac reading (18). In this latter case the lexeme *tableau* has an (old) unusual meaning, referring to the stages of making professional progress.

4) Neither syntactic nor lexical correspondences

A typical example may be:

take the wind out of N's sails / couper l'herbe sous le pied de N

Compare (19) and (20):

19. pas folle la guêpe!

20. a sharp customer!

(19) is a *metalepse*, that is different layers of tropes.

Un guêpier / a wasp's nest is a metaphor meaning *a difficult state of affairs*. When combined with a synecdoche, N₁ becomes une guêpe / a wasp. Finally, an understatement, or litotes, is used in order to express about the animal's clever reaction. 20 seems to be the equivalent of 19, that is N₁ should be fooled in a commercial transaction, but being clever he reacts in a, adequate way. In order to express these subtleties, we have two different images and two different syntactic constructions.

5) No idiomatic correspondence

The expression to take the fifth refers to the US constitution whose fifth amendment allows people not to answer a question if the consequence may incriminate her. There is no equivalent expression in Italian, French, or English.

4. EXPERIMENTS AND MODELLING OF THEIR ACQUISITION IN L2

We have given already in section 3 several examples of SFPs for a restricted domain (deception). Our purpose is to study the SFPs within this domain, and to specify the conditions under which these expressions are used in a given language. In order to do so we will build a structured lexicon. This latter contains conceptual, linguistic and pragmatic information like the position of an expression within a hierarchy, its syntactic constraints, lexical functions in the spirit of Melçuk [MELC 86], communicative function (irony, topicalization) and the surface form of the expression (lexemes).

This study should prove useful in order to discover basic aspects of lexical acquisition in L2. The correspondance of expressions like the ones given in section 3 should provide us with valuable clues concerning the learners' behaviour. We would like to find answers to the following questions:

- is it easier to understand and to produce expressions of type 1 and 2 (the two being quite similar in meaning and form and)?
- what will be prevalent in the choice of an expression, the meaning (image) or the linguistic form?

- are paraphrases a reliable tool to measure the student's comprehension even if she cannot find the exact equivalent in her mother tongue?
- what influence may similarities between two languages (type 2 and 3) have on the learning process (positive/negative interferences)?
- what is the role of correspondences between two languages, what is their respective role compared to the other factors mentioned here above (L1 acquisition)?

In order to answer these questions we have designed transversal and longitudinal experiments for second-language learners of French and English. Having the computer run the experiment will remove the semantic decisions from the intuitions of the researcher and place the burden of proof on the computational mechanisms [KREU 91].

Transversal experiments on recognition and production

Population

Two different transversal experiments will be run in English and French with two types of second-language learner: beginners and students of intermediate level.

Corpus and objectives

We will present various lexical phrases that are syntactically, semantically and pragmatically representative of the different classes. The experiment will be done along the lines of Levorato and Cacciari [LEVO 92]: the idiom under study was the conclusion of an easily understandable, very short story.

Hypotheses

According to Cacciari and Glucksberg [CACC 91] context provides a rich set of clues for interpreting and producing idioms. Given a specific kind learner differences in acquisition depend on the semantic class of the idiom (opaque, transparent or metaphorical meaning). This being so we would like to compare relative ease of idiom acquisition in L1 and L2. We would also like to find out to what extent words frequently used in the mother tongue facilitate the acquisition of these words in a foreign language? In our experiment learners should be able to perform syntactic operation on the idioms. The students' errors should give us valuable information concerning L2 acquisition.

Longitudinal experiments: Lexical awareness among L2 adult learners

In a longitudinal study we want to check and to promote lexical awareness of adult L2 learners. We will build an interactive dictionary that is meant to foster lexical competency of second-language, university level students. They need English as an everyday tool for reading scientific material, and they have to read authentic material (articles) which forms a specific type of "sublanguage". Scientific articles offer a good range of expressions and idioms and they use the same expression over and over.

This environment offers an on-line monolingual dictionary with embedded information, so that only useful information will be displayed. After each reading session it will update the student's personal lexicon by adding the new words and expressions encountered. This kind of lexicon will enable the student to memorize words into their context and to link new contexts to older ones. By the end of two years the student will have encountered most of the words he needs in their appropriate context.

There are basically three ideas behind this tool:

- have the students deduce the word's meaning from context (from the dictionary and from the text itself) and help them to discover the relationship among words;
- let them use the intuition of their native language in order to understand the figures of speech and the meaning of idioms;
- promote free association and cognitive awareness in order to enhance memorization

Another purpose of this approach is to provide adults with a tool that allows them to build a structured lexicon for metaphoric and idiomatic expressions. In fact this second goal will be reached while constraining dictionary and lexicon accesses, which will be carefully tracked and by the management of a personal lexicon by the student. All interactions (lexical access, ...) will be recorded and this trace will provide use with valuable clues concerning the way how literal and metaphorical meanings are understood, and concerning the way how learners link new words to old ones.

The task: at each session the students choose an article of around 800 words and answer a few questions by taking into account the following instructions:

- read the text entirely;
- point out dubious words or expressions;
- check the meaning of unknown words in the dictionary;
- if necessary, consult the personal lexicon (The personal lexicon contains, among other things, the associations a user may have when encountering a given word)

The exercise is done when all questions are answered. If the student does not understand the meaning of a word or expression he is invited to consult the dictionary. At the end of the session students are asked to make "free associations" with the new words. Free associations should enhance ease of word access. All interactions (words checked in the dictionary, free associations,...) are recorded in the personal lexicon.

Computer modelling

Data collected from these experiments should help us to gain a better understanding concerning the learning and production lexical phrases in French and English as a second-language. The next (and to some extent parallel) step will be to define a model of mental lexicon. This model is based on the following three parts:

- 1) a description of lexical phrases with a limited number of concepts, containing all important syntactic, semantic and pragmatic features;
- 2) a model of automatic parsing and generation;

- 3) models of recognition and production strategies for lexical phrases.

The first task can be divided into several subtasks. We will list all words and expressions of a semantic field, in our case, human deception). A structural description of the corresponding SFPs is undertaken in terms of lexicalized TAGs (Tree Adjoining Grammar, for more details see [ABEI 89]).

The same grammar is used in order to describe the syntactic behaviour of SFPs as well as other sentences and to parse all of them. Within the TAG paradigm a SFPs corresponds to a simple entry (that is, a non-compositional entry) which can be made out of several discontinuous items. Insertion of modifiers can easily be handled: it is thus possible to distinguish between an insertion that applies to a sub-component of a SFP, an insertion which modifies the entire expression, or an insertion that rules out the idiomatic reading. Different kinds of transformations can also be handled: passivation, topicalization, cleft constructions, relativization, wh-questions, pronominalization,

The last two hardly ever apply to SFPs. One important point here is to measure the extent to which a SFP deviates syntactically from its literal counterpart, because, if only few of them deviate, we could have a predictive model of SFPs.

SFPs must be described not only in terms of syntax, but also in terms of semantics and pragmatics.

The semantic description can be done in functional terms, that is, in terms that spell out relationships holding between SFPs. The result of such description should make obvious which SFPs are closely related in terms of meaning, what relationship holds between a SFPs and a general concept (N_0 to have N_1 on/ N_0 deceive N_1), or what is the relationship between a SFP and its word components. In order to achieve goal we have to go beyond simple relations such as synonymy and hyponymy. What we need is a set of lexical functions as rich as the ones provided by the Meaning-Text Theory [MELC 86]. The pragmatic description will be based on the most significant parameters governing the use of SFPs in a communicative setting (register, familiarity, frequency, ...).

The four tasks related to part one gives only a prescriptive view of the linguistic knowledge of SFPs in one domain, i.e. the expert/linguist's view. How does that knowledge relate to the learner's mental lexicon in L1 and L2 ? A simplistic way would be to consider the learner's knowledge as subset of the expert's knowledge. This known as the *overlay approach* in learner modelling in Intelligent Tutoring Systems [WENG 87]).

Moreover, the output of this description should give us a classification of SFPs w.r.t. different criteria which can help to structure lexicons in L1 and L2 and predictive model concerning what can be learned most easily. This classification will be tested along the lines presented here above. We believe that this kind of interaction between the empirical work done by psychologists (experiments with people) and the simulation done by computer scientist (description and simulation of a theory) are a fundamental research strategy. Nearly all experiments done in psycholinguistics concerning idioms rely on the designers' linguistic intuitions. This intuition usually is too scarce to allow for building a computational model.

The static part of the mental lexicon (covering L1 and L2) can be divided into roughly two components: declarative, linguistic knowledge and procedural knowledge. Part 2 will account

for the access and production processes, which in the area of natural language processing are called parsing and generation. Important problems which should be addressed are: when parsing an utterance, what factors determine the recognition of several words being an idiomatic expression rather than a sequence of words? (As previously mentioned syntactic information can rule out the idiomatic interpretation, or accept both readings. The point here is to determine what information blocks the literal interpretation. Is it the word's meanings, the co-text?, etc.

With regards to generation, the question is to know what factors determine the selection of an idiom rather than an expression composed of free elements. Systems which let the learner discover these factors empirically have been proposed already in the context of computerized language learning. Having determined the content, SWIM asks the user to try on his own to find the equivalent linguistic structure (words and sentence form). It is only then that the system generates its version [ZOCK 92].

As stressed in section 3, many SFPs are related to tropes, and even if hearers do not need to apply the trope strategy of interpretation in order to recognise them in L1, this strategy is still directly accessible and often used when changing lexical items in expressions [GIBB 90]. In consequence, another important capacity of the parser/generator should be to allow for, or to produce lexical variations. Consider examples (27) and (28), taken from [CRUS 86: 42]:

- 27. They tried to sweeten/sugar the pill.
- 28. They tried to sugar the medicine.

A parser taking (28) as input should be able to access the SFP used in (27) ; conversely, a generator should be able to produce both versions, as they are reasonably close in meaning.

Part 3 refers to the dynamic aspects of the mental lexicon. We use the word "dynamic" in order to refer to the *acquisition* of SFPs rather than to their processing in discourse (analysis/generation). Description of linguistic knowledge (part 1) and the implementation of processes that can access it (part 2) are necessary components for modelling a given knowledge state of the learner. In order to capture the dynamic nature of acquisition we have to represent explicitly the factors that play a role when moving from one (knowledge) state to another. The strategies a second language learner develops in order to understand or produce utterances are an important issue. They can be divided into first or second-language strategies [IRUJ 86], i.e. strategies that apply knowledge either of the source or the target language. A typical example of a target language strategy in generation is the substitution of semantically related words whenever one does not fully remember a SFP. In the same situation the computer would apply the processes (part 2) mentioned here above, in order to preserve the trope.

5. CONCLUSION

Semi-frozen phrases represent a challenge for researchers studying the acquisition of words and expressions in a foreign language: though being frequently used in the mother tongue, their acquisition is considered as being difficult in L2. We have offered a framework in order to study SFPs, to compare them and to explain why they are so difficult for the second language learner. A lot of research has been done in psycholinguistics on the processing of idioms in L1. There have also been some studies on idiom acquisition in L1, but very little work has been done on the processing of idioms and their acquisition in L2. We have reviewed some of the results in L1 that seem to be relevant for L2 learning, and characteristic phenomena of lexical acquisition in a target language such as the correspondence of expressions in the source language, among which tropes may play a crucial role. We have sketched some possible experiments that could be run with L2 learners and a computational model of lexical phrase acquisition. Finally, we have stressed the need for a computational

model in order to move away from the researchers' intuitions and to get at grips with some of the problems that really are at stake in L2 acquisition.

REFERENCES

- [ABEI 89] Abeillé A., Schabes Y.: "Parsing Idioms in Lexicalized TAGs". European Chapter of the ACL, Manchester, April.
- [BOBR 73] Bobrow S., Bell S. (1973): "On catching on to idiomatic expressions". *Memory and Cognition*, 1, pp 343-346.
- [CACC 88] Cacciari C., Tabossi P. (1988): "The comprehension of idioms". *Journal of Memory and Language*, 27, pp 668-683.
- [CACC 89] Cacciari C., Levorato M.C. (1989): "How children understand idioms in discourse". *Journal of Child Language*., 16, pp. 387-405.
- [CACC 91] Cacciari C., Glucksberg S. (1991): "Understanding Idiomatic Expressions: the contribution of Word meanings". In *Understanding Word and Sentence*, Simpson G.B. (ed), Elsevier Science Publishers, North-Holland. pp 217-240.
- [CACC 92] Cacciari C. (1992): "The place of idioms in a literal and metaphorical world. In *Idioms processing, structure and interpretation*., Cacciari C., Tabossi P. (eds), Hillsdale, NJ: Erlbaum.
- [CART 88] Carter R., McCarthy M. (eds) (1988): *Vocabulary and Language Teaching*. Longman.
- [CONN 84] Conenna M. (1984): "Les expressions figées en français et en italien: problèmes lexico-syntaxiques de traduction". *Contrastes*, 10.
- [COWI 83] Cowie A.P., Mackin R., McCaig I.R. (1983): *Oxford Dictionary of Current Idiomatic English*. Volume 1 et 2. Oxford University Press: Oxford.
- [DANL 88] Danlos L. (ed.) (1988) : Les expressions figées. *Langages* , n° 90, juin.
- [D'ELI 90] D'Elia C. (1990): "Idioms dictionaries : Italian and English". *Linguisticae Investigationes*, tome XIV, 2, pp 263-300.
- [FONT 77] Fontanier P. (1977): *Les figures du discours*, Flammarion, Paris, 1ère édition 1818.
- [FREC 85] Freckleton P. (1985): *Une comparaison des expressions de l'anglais et du français*. Doctoral Thesis, Université Paris 7: LADL.
- [GIBB 80] Gibbs W.R. (1980): "Spilling the beans on understanding and memory for idioms in context." *Memory and Cognition*, vol 8, pp.149-156.
- [GIBB 87] Gibbs W.R. (1987): "Linguistic factors in children's understanding of idioms." *Journal of Child Language*, vol 14, pp.569-586.
- [GIBB 89a] Gibbs W.R., Nayak N.P. (1989): "Psycholinguistic Studies on the syntactic behavior of idioms.", *Cognitive Psychology*, 21, pp100-138.
- [GIBB 89b] Gibbs W.R., Nayak N.P., Cutting C. (1989): How to kick the bucket and not decompose: analyzability and idiom processing". *Journal of Memory and Language*, 28, pp 576-593.
- [GIBB 90] Gibbs R.W., O'Brien J.E. (1990): "Idioms and mental imagery: the metaphorical motivation for idiomatic meaning". *Cognition*, 36, pp 35-68.
- [GROS 82] Gross M. (1982): "Une classification des phrases figées du français", *Revue québécoise de linguistique*, Vol 11, n°2, Montréal: Presses de l'Université du Québec à Montréal, pp 151-185.

- [IRUJ 86] Irujo S. (1986): "Don't put your leg in your mouth: transfer in the acquisition of idioms in a second language". *TESOL Quaterly*, vol 20, 2, pp 287-305.
- [LEVO 92] Levorato M.C., Cacciari C (1992): "Children's comprehension and production of idioms: the role of context and familiarity". *Journal of Child Language.*, in press.
- [KREU 91] Kreuz R.J., Graesser A.C. (1991): "Aspects of idiom interpretation: comment on Nayak and Gibbs". *Journal of Experimental Psychology.*, vol. 120, n° 1, pp. 90-92.
- [MANS 83] Manser M. H. (1983): *A dictionary of contemporary idioms*. Pan books: Londres.
- [MELC 86] Mel'Cuk I.(1986): *Dictionnaire Explicatif et Combinatoire du Français Contemporain.*, Presses de l'Université de Montréal, Québec.
- [NATTI 88] Nattinger J. (1988): "Some current trends in vocabulary teaching. *Vocabulary and Language Teaching.*.. Carter R., McCarthy M. (eds). Longman.
- [POPI 88] Popiel S.J., McRae K. (1988): "The figurative and literal senses of idioms, or all idioms are not used equally". *Journal of Psycholinguistic Research*, vol 17, 6, pp 475-487.
- [REY 91] Rey A., Chantreau S. (1991): *Dictionnaire des expressions et locutions*. *Dictionnaires Le Robert*: Paris.
- [SCHW 86] Schweigert W.A. (1986): "The comprehension of familiar and less familiar idioms." *Journal of Psycholinguistic Research*, vol 15, 1, pp.33-45.
- [SWIN 79] Swinney D.A., Cutler A. (1979): "The access and processing of idiomatic expressions." In *Journal of verbal learning and verbal behavior*, n°18, pp 523-534.
- [WEN 87] Wenger E. (1987): *Artificial Intelligence and Tutoring Systems*. Morgan Kaufmann.
- [ZOCK 92] Zock M. (1992): "SWIM or SINK: the problem of communication thought". *The Bridge to International Communication: Intelligent Tutoring Systems for Foreign Language Learning*, Swartz M et Yazdani M. (eds). Springer-Verlag, NATO ASI Series, pp. 235-247.