

UNIVERSITÉ DE GENÈVE

FACULTÉ DES LETTRES

Département de linguistique

Professeur É. Wehrli

Modèle d'analyse lexico-syntaxique des locutions espagnoles

THÈSE

présentée à la Faculté des lettres de l'Université de Genève
pour obtenir le grade de Docteur ès lettres, mention linguistique

par

Jorge Antonio Leoni de León



Thèse N° 650

GENÈVE

2008

La Faculté des lettres, sur le préavis d'une commission composée des MM. les professeurs Jacques MOESCHLER, président du jury ; Éric WEHRLI, directeur de thèse ; Dr. Christopher LAENZLINGER (Université de Genève) ; Rafael RODRÍGUEZ MARÍN (Universidad de Educación Estatal a Distancia, Madrid) et Thierry FONTENELLE (Microsoft Natural Language Group, Redmond, États-Unis) autorise l'impression de la présente thèse, sans exprimer d'opinion sur les propositions qui y sont énoncées.

Genève, le 24 mai, 2008

Thèse N° 650

Le Doyen, Éric WEHRLI

Notes :

Version complète disponible en ligne :

→ <http://www.unige.ch/cyberdocuments/theses2008/LeonideLeonJA/these.pdf>

Informations générales sur le document :

→ <http://www.unige.ch/cyberdocuments/theses2008/LeonideLeonJA/meta.html>

→ <http://archive-ouverte.unige.ch/vital/access/manager/Repository/unige:594>

Téléchargement du document :

→ <http://archive-ouverte.unige.ch/vital/access/services/Download/unige:594/THESIS>

Chapitres

1. Introduction	1
2. Autour des <i>expressions à mots multiples</i>	13
3. Typologie et phénoménologie	72
4. Formalisation lexico-syntaxique	120
5. Le modèle phraséologique	178
6. Conclusions	237

Modèle d'analyse lexico-syntaxique des locutions espagnoles

par

Jorge Antonio Leoni de León

Thèse

présentée à la Faculté des lettres de l'Université de Genève
pour obtenir le grade de Docteur ès lettres, mention linguistique

Mai 2008

Résumé

Dans cette thèse, nous nous intéressons à la représentation informatique de la syntaxe des *expressions à mots multiples* de l'espagnol. Le terme *expressions à mots multiples* désigne un ensemble de phénomènes qui se caractérise par le recours à plusieurs unités lexicales, afin de produire un seul et même effet dénotatif ou stylistique, souvent dans des circonstances où un élément lexical spécifique aurait pu être utilisé. Toutes les catégories grammaticales sont concernées. Par exemple, « ojo de buey » (« œil-de-bœuf ») est un groupe nominal désignant un type de lucarne, tandis que « meter la pata » (« mettre les pieds dans le plat ») est un groupe verbal caractérisant une manière particulièrement maladroite d'intervenir dans une situation.

La syntaxe de ce type d'expressions soulève plusieurs questions liées à leur statut dans le lexique. Cette situation se reflète dans l'ambiguïté terminologique qu'elle entraîne : le niveau du langage désigné par le terme *expressions à mots multiples* n'est pas toujours clair. En effet, cette notion désigne indifféremment des éléments de la parole et de la langue. Dans ce travail, plus précisément dans le chapitre 2, nous distinguons trois concepts : les *vocables*, unités de la parole réalisées comme des séquences de caractères bornées par des espaces blancs, les *lexèmes*, des unités abstraites du lexique, et les *morpholexèmes*, des unités intermédiaires correspondant à un ou plusieurs vocables liés à un seul lexème. Nous réservons la notion de *lemme* aux entrées d'un dictionnaire.

Ces distinctions ont trois conséquences directes : nous rejetons tout d'abord le terme *expressions à mots multiples* en faveur de celui d'unité phraséologique, car notre travail de recherche est de nature lexicographique ; ensuite, nous définissons les unités phraséologiques en termes de liens entre lexèmes ; enfin, ces distinctions nous amènent à établir une typologie des *expressions à mots multiples* sur des nouvelles bases formelles.

Un modèle informatique exigeant de connaître le comportement de l'objet à représenter, nous avons tout d'abord cherché, dans le chapitre 3, à identifier la phénoménologie syntaxique des unités phraséologiques. Nous avons organisé ensuite les résultats dans une typologie comprenant trois grandes classes : les collocations (relations d'alternance lexicale), les fusions thématiques (syntagmes composés d'un verbe support et d'un nom prédicatif) et les unités lexicales plurimembres (les unités phraséologiques proprement dites). Ces derniers comprennent d'une part les lexèmes plurimembres, contenant des expressions nominales, adverbiales et adjectivales, et d'autre part, les phrasèmes, qui sont des expressions verbales. Cette taxonomie comprend encore quelques sous-classes, qui, comme les autres, répondent à des critères lexico-syntaxiques. Notre recherche s'est alors focalisée sur les unités lexicales plurimembres, plus précisément selon l'axe phrasique (paradigmatique, syntagmatique). Les phénomènes paradigmatiques correspondent à des possibilités de substitution dans la structure d'une unité phraséologique, tandis que les phénomènes syntagmatiques impliquent des relations de distance avec comme conséquence une discontinuité de l'expression. Par exemple, sur l'axe syntagmatique, « meter la pata » est une expression verbale qui connaît la cliticisation : « él la metió hasta el fondo » (« il a fait une grosse gaffe »). En revanche, sur l'axe paradigmatique « meter la pata » peut se trouver sous la forme de « meter las de andar ».

Dans le chapitre 4, nous définissons une formalisation lexico-syntaxique, appelée *Tsool*, capable de restituer les caractéristiques lexico-syntaxiques d'une unité phraséologique. *Tsool* est un formalisme riche qui permet, entre autres, l'établissement des relations interlinguistiques à des fins de traduction, de même que la spécification de définitions

et d'exemples, tout en respectant les structures d'arguments. En outre, *Tsool* est conçu pour exploiter des données morpho-syntaxiques depuis un nombre arbitraire de systèmes automatiques de données linguistiques. Dans notre cas, nous exploitons les analyseurs lexical et syntaxique de FIPS, sur lesquels nous basons nos analyses de la phrase. Dans le chapitre 5, notre modèle, nommé \underline{MUL} \overline{KIN} , reproduit la formalisation *Tsool* à l'aide des technologies de l'univers XML traitées avec *Perl*. Nos exemples du système sont basés sur une preuve de concept illustrant les principales caractéristiques du modèle.

En résumé, notre objectif principal est de créer un modèle informatique phraséologique de l'espagnol basé sur la phénoménologie lexico-syntaxique des expressions idiomatiques. Notre recherche aboutit à une classification lexico-syntaxique des unités phraséologiques, classification qui est la base d'une formalisation permettant une automatisation du calcul de l'idiomaticité.

Lexico-syntactic analysis model of Spanish multi-word expressions

by

Jorge Antonio Leoni de León

Thesis

submitted to the Faculty of Humanities of the University of Geneva

in fulfillment of the requirements for the Degree *Docteur ès lettres* in Linguistics

May 2008

Abstract

In this thesis, we focus our attention on the systematic representation of the syntax of Spanish multiword expressions. The term multiword expressions stands for a range of lexical phenomena characterized by the use of several lexical units to produce an equivalent denotative or stylistic effect; frequently, when a single specific lexical element would be sufficient. This phenomenon affects all grammatical categories. For example, the sequence “ojo de buey” is a nominal group that refers to a skylight type, while, “meter la pata” is a verbal group that refers to a particularly clumsy manner of handling a situation.

The syntax of these types of expressions causes several problems associated with their lexical status. Furthermore, there is a terminological ambiguity: the linguistic level assigned to multiword expressions is not always clear. This situation becomes evident because such terms affect both speech and language. In order to better differentiate the lexical elements involved in these phrases, in chapter 2, we have distinguished three concepts: vocables, speech units developed as character sequences limited by blank spaces; lexemes, abstract lexical units; and morpholexemes, intermediate units that belong to one or several vocables related to one single lexeme. The concept lemma (headwork) is limited to the entries in a dictionary. This approach has three direct consequences: first, we dispose of the term multiword expression and adopt in its place the term phraseological

unit, since our investigation is of a lexicographical nature. Second, we define phraseological units as links between lexemes. Finally, the distinction allows us to establish a typology for multiword lexical units upon a new syntactically-oriented formal basis.

Because the proper knowledge of the behavior of a represented object is a requisite in any systematical model, in chapter 3, we identify and describe the syntactic phenomena of phraseological units. Then, we organize the results in a consistent typology of three main types: collocations (lexical alternation relationships), thematic fusions (constituents formed by one support verb and a noun complement), and multimember lexical units (phraseological units themselves). The latter includes, on one hand, multimember lexemes, which contain nominal, adverbial and adjectival expressions, and, on the other hand, phrases that are verbal expressions. Subclasses, also defined according to lexical and syntactic criteria, are included in this taxonomy as well. From this point, our investigation focuses on the multimember lexical units, which are arranged according to the phrase track involved (paradigmatic or syntagmatic). The paradigmatic phenomena result from the possibility of substitution in the structure of the phraseological unit; while the syntagmatic phenomena involve distance relationships from which discontinuity of the expression often occurs. For example, in the syntagmatic track, “meter la pata” is a verbal expression in which the internal argument *mutatis mutandi* can be replaced by a clitic pronoun like: “él la metió hasta el fondo”. Concerning the possibilities in the pragmatic track, the internal argument of that same expression can be replaced by a periphrasis like “meter las de andar”.

In chapter 4, we define a formal notational system called *Tsool*, capable of restoring the lexical and syntactic features of a phraseological unit. *Tsool*, is a versatile instrument that allows users to establish interlinguistic relationships for translation as well as to specify definitions and examples taking the structures of arguments into account. Moreover, *Tsool* was created to represent the use of morphosyntactic data from an arbitrary amount of linguistic data from automated systems. In our case, we use FIPS’ lexical and syntac-

tic parser from the *Laboratoire d'Analyse et de Technologie du Langage* (LATL) of the University of Geneva, upon which we have based our phrase analysis. In chapter 5, our model called MULIN reproduces the *Tsool* formalization through XML technologies processed by several *Perl* applications. Our system examples are based on a proof of concept that displays the main features of the model; as well as our basic processing strategies.

In short, in this investigation we pursue the creation of a systematic Spanish phraseology model based on the lexical and syntactic phenomena of idioms. We establish a formalization of complex lexical unit relationships that allows the calculation of an automated idiomatic nature from a typology that takes into account the behavior of phraseological units at a sentence level.

Modelo de análisis léxico-sintáctico de las locuciones españolas

de

Jorge Antonio Leoni de León

Tesis

presentada en la Facultad de Letras de la Universidad de Ginebra
para obtener el grado de Doctor en Letras con mención en Lingüística

Mayo 2008

Resumen

En esta tesis nos interesamos en la representación informática de la sintaxis de las unidades léxicas pluriverbales del español. El término francés, *expressions à mots multiples* designa un conjunto de fenómenos léxicos caracterizados todos por recurrir al empleo de varias unidades léxicas con el fin de producir un mismo efecto denotativo o estilístico, a menudo en circunstancias en las que un solo elemento léxico específico habría sido suficiente. Todas las categorías gramaticales se ven afectadas por esta fenomenología. Por ejemplo, la secuencia “ojo de buey” es un grupo nominal que designa un tipo de tragaluz, mientras que “meter la pata” es un grupo verbal que se refiere a una manera particularmente torpe de intervenir en una situación.

La sintaxis de este tipo de expresiones plantea varios problemas vinculados con su estatus en el léxico. A lo que se agrega una situación terminológicamente ambigua: el nivel del lenguaje designado por el término francés *expressions à mots multiples* no está siempre claro. Esta situación se manifiesta en el hecho de que esta noción designa tanto elementos del habla, como de la lengua. Con el fin de diferenciar mejor los elementos léxicos implicados en la frase, en el capítulo 2 distinguimos tres conceptos : los *vocablos*, unidades del habla que se realizan como secuencias de caracteres limitadas por espacios en blanco, los *lexemas*, unidades abstractas del léxico, y los *morfolexemas*, unidades intermedias que corresponden a uno o varios vocablos relacionados con un solo lexema.

La noción de *lema* queda limitada a las entradas de un diccionario. Estas distinciones tienen tres consecuencias directas: en primer lugar descartamos el término *expressions à mots multiples*, en su lugar adoptamos el término *unidad fraseológica*, dado que nuestra investigación es de naturaleza lexicográfica; luego, definimos las unidades fraseológicas en términos de vínculos entre lexemas; finalmente, estas distinciones nos llevan a establecer una tipología de las unidades léxicas plurimembres sobre nuevas bases formales de orientación sintáctica.

Siendo que el conocimiento adecuado del comportamiento del objeto por representar constituye una de las exigencias de todo modelo informático, en el capítulo 3 identificamos y describimos la fenomenología sintáctica de las unidades fraseológicas. Luego organizamos los resultados en una tipología consistente en tres grandes clases: las colocaciones (relaciones de alternancia léxica), las fusiones temáticas (sintagmas compuestos de un verbo soporte y de un sustantivo predicativo) y las unidades léxicas plurimembres (las unidades fraseológicas propiamente dichas). Estas últimas incluyen, por un lado, los lexemas plurimembres, que contienen las expresiones nominales, adverbiales y adjetivas, y, por otra parte, los frasemas que son expresiones verbales. En esta taxonomía también incluimos otras subclases, definidas de igual manera según criterios léxico-sintácticos. A partir de ese momento, nuestra investigación se focalizó en las unidades léxicas plurimembres, las que abordamos según el eje frástico implicado (paradigmático o sintagmático). Los fenómenos paradigmáticos corresponden a posibilidades de sustitución en la estructura de la unidad fraseológica, mientras que los fenómenos sintagmáticos implican relaciones de distancia, de las que resulta muchas veces una discontinuidad de la expresión. Por ejemplo, en el eje sintagmático, “meter la pata” es una expresión verbal cuyo argumento interno *mutatis mutandi* puede ser reemplazado por un pronombre clítico: “él la metió hasta el fondo”. En lo concerniente a sus posibilidades en el eje paradigmático, el argumento interno de esa misma expresión puede ser sustituido por una perífrasis: “meter las de andar”.

En el capítulo 4, definimos un sistema de notación formal, llamado *Tsool*, capaz de restituir las características léxico-sintácticas de una unidad fraseológica. *Tsool* es un formalismo versátil que permite, entre otros, tanto establecer relaciones interlingüísticas para la traducción, como especificar definiciones y ejemplos, respetando siempre las estructuras de argumentos. Además, *Tsool* ha sido concebido para representar la explotación de datos morfosintácticos de una cantidad arbitraria de sistemas automatizados de datos lingüísticos. En nuestro caso, nosotros recurrimos a los analizadores léxico y sintáctico de FIPS del *Laboratoire d'Analyse et de Technologie du Langage* (LATL) de la Universidad de Ginebra, sobre el cual basamos nuestros análisis de la frase. En el capítulo 5, nuestro modelo, llamado MUKIN, reproduce la formalización *Tsool* por medio de tecnologías del universo XML procesadas por medio de un conjunto de programas escritos en *Perl*. Nuestros ejemplos del sistema están basados en una prueba de concepto que ilustra las principales características del modelo, así como nuestras estrategias básicas de procesamiento.

En resumen, en esta investigación buscamos crear un modelo informático fraseológico del español basado en la fenomenología léxico-sintáctica de las expresiones idiomáticas. A partir de una tipología que toma en cuenta el comportamiento de las unidades fraseológicas al nivel de la oración, en este trabajo establecemos una formalización relacional de las unidades léxicas complejas que permite automatizar el cálculo de idiomática.

Modelo pri leksik-sintaksa analizo de la multvortaj unuoj en la hispana lingvo

verkita de

Jorge Antonio Leoni de León

Tezo

prezentita en la Fakultato pri Beletroj de la Universitato de Ĝenevo

por atingi la gradon de Doktoro pri Beletro, branĉo pri Lingvistiko

Majo 2008

Resumo

En tiu ĉi tezo ni interesiĝas pri la komputika sintaksorepresentado de la multvortaj esprimoj de la hispana lingvo. La termino multvortaj esprimoj aludas al leksika fenomenaro karakterizita per la uzo de pluraj leksikunuoj kun la celo produkti saman denotaciivan aŭ stilan efekton, ofte en cirkunstancoj en kiuj ununura leksika elemento sufiĉus. Tiu ĉi fenomenaro afektas ĉiujn gramatikajn kategoriojn. Ekzemple, la sekvenco “ojo de buey” (“bovokulo”) estas substantiva grupo indikanta ian fenestron; krome “meter la pata” (“fari miŝpaŝon”) estas verba grupo referanta al speciale mallerta maniero interveni en situacio.

La sintakso de tiaj ĉi esprimoj starigas plurajn problemojn ligitajn al ilia statuso en la leksiko. Krome, ankaŭ ekzistas terminologie ambigua situacio : la lingvonivelo aludita per la termino multvortaj esprimoj ne ĉiam estas klara. Tiu ĉi situacio manifestiĝas en la fakto que tiu ĉi nocio indikas elementojn apartenantajn, jen al la parolo, jen al la lingvo. Cele pli efikan diferencigon de la leksikaj elementoj de la frazo, en la ĉapitro 2 ni distingas tri konceptojn : la *leksemaĵoj*, parolunuoj, kiuj realiĝas kiel karaktersekvencoj troviĝantaj inter blankaj spacoj, la *leksemoj*, abstraktaj unuoj de la leksiko, kaj la *morfoleksemoj*, mezaj unuoj korespondantaj ĉu al unu, ĉu al pluraj *leksemaĵoj* ligitaj al unusola leksemo. Ni rezervas la nocion *lemo* al leksikonaj ŝlosilvortoj. Tiuj ĉi distingo havas tri rektajn konsekvencojn : unue, ni forĵetas la terminon “multvortaj esprimoj”, adoptante la pli precizan terminon “frazeologian unuon”, ĉar la naturo de nia esploro estas leksikografia;

poste, ni difinas la frazeologiajn unuojn kiel ligojn inter leksemoj; fine, tiuj ĉi distingoj kondukas nin al establo de tipologio de la leksikaj plurmembraĵoj sur formalaj bazoj sintakse orientitaj.

Ĉar adekvata kono pri la konduto de la reprezentota objekto konstituas unu el la postuloj de ajna komputika modelo, en la ĉapitro 3 ni identigas kaj priskribas la sintaksan fenomenologion de la frazeologiaj unuoj. Poste, ni organizas la rezultojn en tipologio konsistanta el tri grandaj klasoj : apudmetoj (leksikaj alternrilatoj), temaj kunfandoj (sintagmoj kunmetitaj el subtena verbo kaj predikativa substantivo) kaj plurmembraĵoj leksikaj unuoj (la frazeologiaj unuoj mem). Ĉi lastaj inkludas, unuflanke, la plurmembraĵojn leksemojn, kiuj enhavas substantivajn, adverbajn kaj adjektivajn esprimojn, kaj, aliflanke, la frazemojn, kiuj estas verbaj esprimoj. En tiu ĉi taksonomio, ni ankaŭ inkludas aliajn subklasojn, difinitajn laŭ ekvivalentaj leksik-sintaksaj kriterioj. De tiu momento, nia esploro fokusiĝis pri la plurmembraĵoj leksikaj unuoj, kiujn ni analizis laŭ fraza akso (ĉu paradigma, ĉu sintagma). La paradigmaj fenomenoj korespondas al anstataŭigebleco de elementoj apartenantaj al la strukturo de la frazeologia unuo; sintagmaj fenomenoj implicas distancrilatojn, el kiuj ofte rezultas diskontinueco de la esprimo. Ekzemple, en la sintagma akso, la frazeologia unuo “meter la pata” estas verba esprimo, kies interna argumento *mutatis mutandi* povas esti anstataŭigita per klitika pronomo : “él la metió hasta el fondo”. Koncerne ĝiajn eblojn en la paradigma akso, la interna argumento de tiu sama esprimo povas esti anstataŭigita per ĉirkaŭfrazo : “meter las de andar”.

En la ĉapitro 4, ni difinas formalan notacisistemon, nomitan *Tsool*, kapablan reprodukti leksik-sintaksajn karakterizojn de frazeologia unuo. *Tsool* estas fleksebla formalismo permesanta, inter aliaj, establون de interlingvaj rilatoj cele tradukadon kaj specifigon de difinoj kaj ekzemploj, respektante ĉiam la argumentstrukturo. Krome, *Tsool* estas konceptita por reprezenti la ekspluaton de morfosintaksaj datumoj el ajna nombro de aŭtomatigitaj sistemoj de lingvaj datumoj. En nia kazo, ni utiligis la leksikan kaj sintaksan analizilojn de FIPS de la *Laboratoire d'Analyse et de Technologie du Langage* (LATL)

de la Universitato de Ĝenevo, sur kiu ni bazigis niajn frazanalizojn. En la ĉapitro 5, nia modelo, nomita *MUKIN*, reproduktas la formalismon *Tsool* per teknologioj de la universo XML procesitaj per programaro verkita en *Perl*. Niaj ekzemploj pri la sistemo estas bazitaj sur konceptpruvo ilustranta la ĉefajn karakterizojn de la modelo, same kiel niajn bazajn strategiojn de procesado.

Resume, en tiu ĉi esploro ni celis krei frazeologian komputikan modelon de la hispana surbaze de la leksik-sintaksa fenomenaro de idiotismaj esprimoj. Nia esplorkonkludo, el leksik-sintaksa tipologio de frazeologiaj unuoj, establas rilatecan formalismon de kompleksaj leksikaj unuoj permetantan aŭtomatigi kalkulon pri idiotismeco.

Modello d'analisi lessico-sintattica delle unità polirematiche in spagnolo

di

Jorge Antonio Leoni de León

Tesi

presentata alla Facoltà di Lettere dell'Università di Ginevra

per ottenere il titolo di *Docteur ès lettres* in Linguistica

Maggio 2008

Riassunto

Oggetto di questa tesi è la rappresentazione informatica della sintassi delle *unità polirematiche* in spagnolo. Il termine *unità polirematica* designa un insieme di fenomeni caratterizzati dal ricorso a diverse unità lessicali, con lo scopo di produrre un unico effetto denotativo o stilistico, spesso in situazioni in cui un elemento lessicale specifico avrebbe potuto essere utilizzato. Tutte le categorie grammaticali ne sono interessate. Per esempio, « ojo de buey » (« occhio di bue ») è un gruppo nominale che designa un tipo di lucernario, mentre « meter la pata » (« fare una gaffe ») è un gruppo verbale che caratterizza un modo inopportuno di intervenire in una situazione.

La sintassi di questo tipo di espressioni solleva diverse questioni relative al loro statuto nel lessico. Tale situazione si riflette nell'ambiguità terminologica che ne deriva: il livello di linguaggio designato dal termine *unità polirematica* non sempre è chiaro. Questa nozione, infatti, designa indifferentemente elementi della parola e della lingua. In questo lavoro, e più precisamente nel capitolo 2, distinguiamo tre concetti: i *vocaboli*, unità di parola realizzate come sequenze di caratteri limitate da spazi bianchi, i *lessemi*, unità astratte del lessico, e i *morfolessemi*, unità intermedie corrispondenti a uno o a più vocaboli legati ad un solo lessema. Conserviamo la nozione di *lemma* per le voci di un dizionario. Queste distinzioni hanno tre conseguenze dirette: rigettiamo prima di tutto il termine *unità polirematica* in favore di quello di unità fraseologica, dal momento che il

nostro lavoro di ricerca è di natura lessicografica; definiamo, in secondo luogo, le unità fraseologiche in termini di legami tra lessemi; queste distinzioni ci portano, quindi, a stabilire una tipologia di *unità polirematiche* su nuove basi formali.

Data l'esigenza del modello informatico di conoscere il comportamento dell'oggetto da rappresentare, abbiamo, in primo luogo, cercato di identificare, nel capitolo 3, la fenomenologia sintattica delle unità fraseologiche. Abbiamo poi organizzato i risultati in una tipologia comprendente tre grandi classi: le collocazioni (relazioni di alternanza lessicale), le fusioni tematiche (sintagmi composti da un verbo supporto e da un nome predicativo) e le unità lessicali composte (le unità fraseologiche propriamente dette). Queste ultime comprendono da una parte i lessemi composti, contenenti espressioni nominali, avverbiali e aggettivali, e dall'altra parte i frasemi, espressioni verbali. Tale tassonomia comprende ancora alcune sotto classi, che, come le altre, rispondono a criteri lessico-sintattici. La nostra ricerca si è dunque focalizzata sulle unità lessicali composte, soprattutto sull'asse frastico (paradigmatico, sintagmatico). I fenomeni paradigmatici corrispondono a possibilità di sostituzione nella struttura di una unità fraseologica, mentre i fenomeni sintagmatici implicano relazioni di distanza con conseguente discontinuità dell'espressione. Per esempio, sull'asse sintagmatico, « meter la pata » è una espressione verbale che conosce la cliticizzazione, « él la metió hasta el fondo », mentre sull'asse paradigmatico « meter la pata » può trovarsi sotto la forma « meter las de andar ».

Nel capitolo 4, definiamo una formalizzazione lessico-sintattica, chiamata *Tsool*, capace di restituire le caratteristiche lessico-sintattiche di una unità fraseologica. *Tsool* è un formalismo ricco che permette, tra le altre cose, di stabilire relazioni interlinguistiche a fini traduttivi, come pure di specificare definizioni ed esempi, il tutto nel rispetto delle strutture d'argomenti. Inoltre, *Tsool* è concepito per sfruttare dati morfosintattici da un numero arbitrario di sistemi automatici di dati linguistici. Nel nostro caso, utilizziamo gli analizzatori lessicale e sintattico di FIPS del Laboratorio di Analisi e Tecnologia del Linguaggio (LATL) dell'Università di Ginevra, sui quali basiamo le analisi della frase.

Nel capitolo 5, il nostro modello, chiamato MULIN, riproduce la formalizzazione *Tsool* con l'aiuto di tecnologie dell'universo XML trattate con *Perl*. Gli esempi di sistema sono basati su una prova di concetto che illustra le principali caratteristiche del modello.

Riassumendo, il nostro obiettivo principale è creare un modello informatico fraseologico dello spagnolo basato sulla fenomenologia lessico-sintattica delle espressioni idiomatiche. La nostra ricerca porta ad una classificazione lessico-sintattica delle unità fraseologiche che è alla base di una formalizzazione in grado di automatizzare il calcolo dell'idiomaticità.