A corpus-based approach to argument structure

José M. García-Miguel

Universidade de Vigo
gallego@uvigo.es

2007 RRG International Conference (México)

Goals

• To present some possibilities offered by a syntacto-semantic database (ADESSE) for the study of the interactions of verbs and constructions in Spanish
• To introduce the main criteria used in the building of such a database
• To overview some general features of argument structure in Spanish
• To compare both the approach and the results of ADESSE with some insightful proposals of RRG theory

Functional Grammar(s)

• This talk is not specifically about RRG, but it takes as background many ideas shared by most functionalists

• Functionalists share the idea that grammar associates forms with meanings and discourse functions:
  – language is a “system of communicative social action in which grammatical structures are employed to express meaning in context” (Van Valin 2005: 1)

• They differ with respect to (among other things):
  – Standards of adequacy (typological, psychological, ...)
  – How the conceive the relation system - use
  – The role of formalization and the specifics of the formalism
   Moderate functionalists (RRG) ↔ Extreme functionalists

(A simplified model of)
Functional grammar

CONSTRUCTIONAL SCHEMATA

LANGUAGE USE

LEXICON

‘entrenchment’
(frequency → memory)

Variation and Grammar: the 'emergentist' view

• "Grammar is built up from specific instances of use which marry lexical items with constructions; it is routinized and entrenched by repetition and schematized by the categorization of exemplars" (Bybee 2006)

• "Grammar is not fixed and absolute with a little variation sprinkled on the top, but it is variable and probabilistic to its very core" (Bybee & Hopper 2001)

Corpus linguistics

Nowadays, to study language use and frequency of words and constructions means corpus linguistics

Computers have made possible the quick search of large bodies of real text and make easier the task of analysing, annotating and storing linguistic data

Some linguists (e.g. Butler) think that functional linguistics should be not only corpus-based but corpus-driven
Corpus linguistics

Some problems:
- The words and patterns that we find in the corpus should not be confused with the words and patterns that are possible in the language.
- A corpus cannot tell us what is not possible.
- Every fragment in the corpus needs some analysis and interpretation by the linguist.

"The conclusion is that 'intuition-based' linguists and 'corpus-based' linguists need each other. Or better, that the two kinds of linguists, wherever possible, should exist in the same body" (Fillmore 1992: 35)

Corpus linguistics and syntax

"Corpus linguistic research has been largely limited to phenomena that can be accessed via searches on particular words (…) However, a (theoretical) syntactician is usually interested in more abstract structural properties that cannot be investigated easily in this way" (Manning 2003: 294)

I.e., it is not always enough with google search of raw text, nor even with morphosyntactically annotated corpora (or with texts accompanied of interlinearized glosses).

⇒ We need detailed syntactic and semantic annotation of corpora.

ADESSE

ADESSE =
Base de datos de verbos, Alternancias de Diátesis y Esquemas Sintactico-Semánticos del Español
[Syntactic Database of Verbs, Diathesis Alternations and Constructional Schemas of Spanish]
http://adesse.uvigo.es/

- An on-going project funded by Spanish MEC and EU funds

Goal
- A database with syntactic and semantic information for all the verbs and clauses in a corpus of Spanish

ADESSE antecedent

- BDS
Base de Datos Sintácticos del español actual
http://www.bds.usc.es/

- A database with the (manual) syntactic analysis of 159,000 clauses of the corpus ARThUS
- ARTHUS Corpus (Archivo de Textos Hispánicos de la Universidad de Santiago de Compostela)
  - 1.5 million words
  - Textual genres: narrative (37%), spoken (19%), essay (17%), theater (15%), journalistic (12%)
  - Origin: Spain (79%), Americas (21%)

BDS ⇔ ADESSE


Syntactic information
Grammatical features of clauses, verbs and arguments of the corpus
[Each record (clause): 64 fields]

- 159,000 clauses
- 3,500 verbs
- 13,500 valency patterns

[Univ. of Vigo 2002-]

Semantic information
- Verb senses
- Verb classes
- Semantic roles

ADESSE

REGALAR
| Acep. | 1.0 Obtener, dar algo de forma gratuita |
| Class.: | [222] Transferencia |
| Activa. | S |
| Arg.: | 3 |

ARGUMENTOS
- ORDEN
- NIVEL
- CATEGORIAS

para recibir a los recién casados. Al servirle el
explicarle un auténtico: conversar con su nombre
grabado en letras góticas bajo el escudo de la fábrica.
(part of) a record in ADESSE

<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>TEXT</td>
<td>Al novio le regalaron un automóvil convertible [CRO: 44, 1]</td>
</tr>
<tr>
<td>PRED</td>
<td>REGALAR</td>
</tr>
<tr>
<td>Verb class</td>
<td>Transfer of possession</td>
</tr>
<tr>
<td>Voice</td>
<td>Active</td>
</tr>
<tr>
<td>A0</td>
<td>le</td>
</tr>
<tr>
<td>A1</td>
<td>automóvil</td>
</tr>
<tr>
<td>A2</td>
<td>novio</td>
</tr>
<tr>
<td>Agreement</td>
<td>3pl</td>
</tr>
<tr>
<td>Synt Role</td>
<td>Donor: Possession, Receiver</td>
</tr>
<tr>
<td>Synt Category</td>
<td>NP</td>
</tr>
<tr>
<td>Synt Schema</td>
<td>S D I</td>
</tr>
<tr>
<td>Order</td>
<td>IVD</td>
</tr>
</tbody>
</table>

BDS/ADESSE

Other grammatical features
- Clause:
  - Clause Type (main, subordinate, …)
  - Mood
  - Tense
  - Modal and Phase Auxiliaries
  - Negation
  - Illocutionary force
  - Voice
  - …
- Arguments:
  - Definiteness
  - Number
  - Person
  - …

RRG representation and ADESSE features

BDS/ADESSE database aims to be theory-neutral
- it only assumes common Basic Linguistic Theory (in the sense proposed by Dixon)
- but is fairly compatible with functional and constructional grammars
- the approach is aimed to correct or complement basic linguistic theory (or theories) in the light of corpus evidence

Basic strategies: Verbs and arguments in ADESSE
- BDS provides a syntactic characterization of arguments and constructions
  - ESCRIBIR 'write'
    - Subj – DO – IO
    - Subj – DO
    - ...
  - SUSTITUIR 'substitute, replace'
    - Subj – DO – por NP
    - Subj – DO
    - ...
  - ENSEÑAR
    - Subj – DO – IO
    - Subj – DO – a Inf

Verbs and arguments in ADESSE
In many cases, each syntactic construction selections a subset of the potential participants of the scene evoked by the verb
b) Juan [0] escribió una carta [1] 'John wrote a letter'
c) Juan [0] le escribió a su madre [2] 'John wrote to his mother'

The task in ADESSE is to annotate which of the potential participants is selected in each syntactic schema
Verbs and arguments in ADESSE

The same syntactic construction can be mapped with different configurations of semantic arguments

- **Sustituir** 'replace'
  ‘Deco replaced Xavi’
  b) Rijkaard [0] sustituyó a Xavi [2]
  ‘Rijkaard replaced Xavi with Deco’

The same set of semantic arguments can be linked to different syntactic patterns

- **Enseñar** ‘teach’
  ‘She taught him her language’
  ‘She taught the baby how to walk’

Verbs and Arguments in ADESSE

- **Valency potential** of a lexical entry:
  - which arguments can be selected by a given verb?

- **Valency realizations** (diatheses):
  - which arguments are actually expressed
  - which is syntactic realization of each argument
  - voice

(The strategy in ADESSE is to define the valency potential of each verb entry and to register in the corpus all the valency realizations)

Arguments and gradience

- Valency patterns occur in the corpus with different frequency, ranging from the more usual to the rare and unexpected.

- As a consequence, verb arguments are not always syntactically realized
  - Obligatoriness – optionality of arguments is not a yes or no matter, but a gradient
  - Obligatory arguments are those (referential) elements more frequently tied to the verb in texts

- Because obligatoriness is one of the main criteria for the argument – adjunct distinction, this is also a gradient

Arguments of *Enseñar* ‘teach’ (N clauses = 139)

<table>
<thead>
<tr>
<th></th>
<th>Teacher</th>
<th>Learner</th>
<th>Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>138</td>
<td>99.3%</td>
<td>79.9%</td>
</tr>
<tr>
<td>1</td>
<td>111</td>
<td>79.9%</td>
<td>99.9%</td>
</tr>
<tr>
<td>2</td>
<td>105</td>
<td>75.5%</td>
<td>79.9%</td>
</tr>
</tbody>
</table>

Arguments of *Escribir* ‘write’ (N clauses = 321)

<table>
<thead>
<tr>
<th></th>
<th>Writer</th>
<th>Text</th>
<th>Receiver</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>300</td>
<td>64.8%</td>
<td>26.5%</td>
<td>3.1%</td>
</tr>
<tr>
<td>1</td>
<td>208</td>
<td>64.8%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>85</td>
<td>26.5%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>10</td>
<td>3.1%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Frequency and argument structure

- ‘Argument structure’ needs to be replaced by a greatly enriched probabilistic theory capturing the entire range of combinations of predicates and participants that people have stored as sorted and organized memories of what they have heard and repeated over a lifetime of language use (Thompson & Hopper 2001: 47)

Frequency and argument structure

Manning (2003)
- Many subcategorization distinctions presented in the linguistics literature as categorical are actually counterexemplified in studies of large corpora of written language use.
- We can get a much better picture of what is going on by estimating a probability mass function (pmf) over the subcategorization patterns for the verbs in question.
- We can put a probability function over the kinds of dependents to expect with a verb or class, conditioned on various features. (302)

Frequency and argument structure

Proposals
- The meaning of a verb determines its contexts of use and is determined by its contexts of use
- Argument structure is a generalization over registered use, where frequency/probability of cooccurrence is a specially important factor in the entrenchment of a valency pattern.
- Instead of obligatory arguments, or participants inherent to a scene, our past linguistic experiences provides us with certain probability expectations about the reference to a certain participant type in the scenes evoked by a verb

Verbs and constructions

A Basic problem: Polisemy and contextual accommodation. Changes in meaning when a verb enters alternating valency patterns
- The formalization of syntactic alternations
  - different lexical entries: each different meaning is a different verb entry
  - lexical rules (RRG), relating different LS
  - underspecification: only one verb meaning, the differences in meaning should be attributed to the constructions

Verbs and constructions

- Strategies in ADESSE
  ¿lexical rules or constructions?
  - Underspecification: Reduce lexical entries to a minimum (searching wider coverage of the corpus and less task consuming)
  - (new strategy in ADESSE-II): Levels of granularity in the definition of verb senses

Lexical entries in ADESSE

Two levels:
- Level 1: Macro-aception 'verb meaning', associated with a semantic domain and a set of participant roles
- Level 2: (Sub)aception 'verb senses' [work in progress]

<table>
<thead>
<tr>
<th>LEVEL 1</th>
<th>LEVEL 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conocer-1 'know'</td>
<td>Conocer 1.1 'get to know'</td>
</tr>
<tr>
<td></td>
<td>Conocer 1.2 'recognize, distinguish'</td>
</tr>
<tr>
<td></td>
<td>Conocer 1.3 'understand, know deeply'</td>
</tr>
<tr>
<td>Enseñar-1 'show'</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Enseñar-2 'teach'</td>
</tr>
<tr>
<td>Volver-1 'return, go back'</td>
<td>Volver 1.1 (lit) 'return to a place'</td>
</tr>
<tr>
<td></td>
<td>Volver 1.2 (met) 'return to a state or activity'</td>
</tr>
<tr>
<td>Volver-2 'turn round'</td>
<td></td>
</tr>
<tr>
<td>Volver-3 'cause to become'</td>
<td></td>
</tr>
</tbody>
</table>
Verb Entries

- (At level 1) We distinguish verb entries when they are associated with different sets of semantic roles – that means differences in ‘valency potential’, not in ‘valency realization’.
- We try to limit sense distinctions to a minimum, but we must distinguish senses that cannot be ‘unified’:
  - partir 1 (‘go away’) vs. partir 2 (‘break’)
  - saber 1 (‘know’) vs. saber 2 (‘taste’)
- or (less clearly) senses which are related with different lexical classes
  - enseñar 1 (‘show’) vs enseñar 2 (‘teach’)
- Anyway, there no clear boundaries between senses at any level (cf. Kilgarriff 1997)

Unifying verb senses

Typically included in one single verb entry (level 1):

- Diathesis alternations (causative / inchoative, locative alternation, …)
  - Semantic differences are attributed primarily to the construction, not to the verb
- Paradigmatic alternatives within an argument (For ex: write a letter / a novel / a musical work)
  - Meaning accommodation or co-compositionality, but not different verb senses
- Metaphoric and other figurative uses
  - they are annotated as the literal uses, but marked as figurative

Paths of Schematization

![Diagram](image)

Syntactic Patterns ↔ Verbs

Once we have defined verb entries and syntactic patterns, we can take two complementary (not necessarily incompatible) points of view concerning the association of verbs and constructions:

- P. of v. of the verb: alternations (for ex., Levin 1993) of valency patterns keeping, as far as possible, the lexical elements
  - Me enseñó a cantar “She taught me to sing”
  - Me enseñó canto “She taught me singing”
- P. de v. of the construction: ‘surface generalizations’ concerning uses of a constructional schema with different lexical elements (Goldberg 1995, 2000; also Dowty 1998)
  - Me enseñó a cantar “She taught me to sing”
  - Me obligó a cantar “She forced me to sing”

Verbs and alternations

- Many lexicalist approaches have focused on whether verbs admits a certain alternation, for ex. the causative alternation
- (or, alternatively, whether they can be combined with a certain type of argument)

<table>
<thead>
<tr>
<th>Verb</th>
<th>A1</th>
<th>A0 – A1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cambiar</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Crecer</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Aprender</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Enseñar</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

but this is not a yes/no question, although the syntagmatic axis can serve to describe the behavioral profile of verbs

Arguments and behaviour profile: Change of state verbs

![Graph](image)
The meaning of a constructional schema is established
that is, any constructional schema (vgr., passive, double
We can see the relation between verbs and constructions
particular utterances that instantiate the schema
and learned) by generalizing from the meaning of
from any other alternating schema.

The meaning of a constructional schema comes from its association with a class of verbs.
The strength of the association of a constructional schema and a
verb should be measured on the basis of a (syntactically annotated)
corpus
• It is hypothesized that the differences in frequency are
motivated by differences in meaning

Arguments and behavior profile: Verbs of knowledge

"Surface generalizations"

We can see the relation between verbs and constructions
from the point of view of the constructions.
• That is, any constructional schema (vgr., passive, double
object, …) can be described by itself and not as derived
from any other alternating schema.
• The meaning of a constructional schema is established
and learned) by generalizing from the meaning of
particular utterances that instantiate the schema
– An essential part of the characterization of a constructional schema
– The strength of the association of a constructional schema and a
verb should be measured on the basis of a (syntactically annotated)
corpus

Valency alternations and behavior profile

• Each lexical element can be described in terms of the features and contructions it combines with in context (it syntagmatic profile)
• A relevant part of the profile of a verb are its constructional schemas, the realizations of its arguments, and the frequencies of schemas and argument realizations
• Two verbs of the same class may have similar syntagmatic combination and differ in the relative frequency of each combination, and as a consequence in the relative frequency of their core arguments

Verbs in the schema <Subj DO IO>

<table>
<thead>
<tr>
<th>VERB</th>
<th>Meaning</th>
<th>N.</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>DAR</td>
<td>'Give'</td>
<td>1272</td>
<td>Mi padre sólo me da dinero para estudiar [SON: 115]</td>
</tr>
<tr>
<td>DECIR</td>
<td>'Say, tell'</td>
<td>599</td>
<td>Le dije que no quería volver [TER: 048]</td>
</tr>
<tr>
<td>HACER</td>
<td>'Make'</td>
<td>514</td>
<td>Le hizo prometer de llevarlo al local [TER: 119]</td>
</tr>
<tr>
<td>CONTAR</td>
<td>'Tell'</td>
<td>308</td>
<td>Rezole lo contaba que el otro (iba) siempre dando [SON: 223]</td>
</tr>
<tr>
<td>PEDIR</td>
<td>'Ask, request'</td>
<td>273</td>
<td>No podía lo esperar, no [BAI: 413]</td>
</tr>
<tr>
<td>PREGUNTAR</td>
<td>'Ask, inquire'</td>
<td>218</td>
<td>Me preguntó si tenía dinero [LAB: 288]</td>
</tr>
<tr>
<td>PERMITIR</td>
<td>'Allow'</td>
<td>124</td>
<td>¿Me permite su teléfono? [SON: 285]</td>
</tr>
<tr>
<td>OFRECER</td>
<td>'Offer'</td>
<td>119</td>
<td>No puedo ofrecerla grandes cualidades [LAB: 115]</td>
</tr>
<tr>
<td>EXPlicAR</td>
<td>'Explain'</td>
<td>112</td>
<td>Lo explicó como funcionaba la impresora [TER: 062]</td>
</tr>
<tr>
<td>PONER</td>
<td>'Put'</td>
<td>101</td>
<td>Si se lo uso un día le pedirá otro [BAI: 104]</td>
</tr>
<tr>
<td>TRAER</td>
<td>'Bring'</td>
<td>98</td>
<td>Me trae un kilo de bombones a mamá [HIS: 094]</td>
</tr>
<tr>
<td>DEJAR</td>
<td>'Leave'</td>
<td>96</td>
<td>El profesor le dejaba la casa para que la habitara [HIS: 149]</td>
</tr>
</tbody>
</table>

Verb classes of clauses in the schema Active <Subj DO IO>

<table>
<thead>
<tr>
<th>CLASS</th>
<th>Clauses</th>
<th>Verbs</th>
<th>Example verbs</th>
</tr>
</thead>
<tbody>
<tr>
<td>22</td>
<td>2568</td>
<td>80</td>
<td>dar 'give'</td>
</tr>
<tr>
<td>4</td>
<td>2419</td>
<td>108</td>
<td>decir 'say, tell'</td>
</tr>
<tr>
<td>6</td>
<td>806</td>
<td>18</td>
<td>hacer 'do, make'</td>
</tr>
<tr>
<td>31</td>
<td>686</td>
<td>88</td>
<td>traer 'bring'</td>
</tr>
<tr>
<td>1</td>
<td>615</td>
<td>64</td>
<td>recordar 'remember'</td>
</tr>
<tr>
<td>32</td>
<td>430</td>
<td>140</td>
<td>abrir 'open'</td>
</tr>
<tr>
<td>33</td>
<td>305</td>
<td>80</td>
<td>tocar 'touch'</td>
</tr>
<tr>
<td>21</td>
<td>180</td>
<td>17</td>
<td>costar 'cost'</td>
</tr>
<tr>
<td>5</td>
<td>118</td>
<td>20</td>
<td>causar 'cause'</td>
</tr>
<tr>
<td></td>
<td>8127</td>
<td>615</td>
<td></td>
</tr>
</tbody>
</table>

S D/lan a Inf

<table>
<thead>
<tr>
<th>VERB</th>
<th>Class</th>
<th>N</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>OBLIGAR</td>
<td>Obligación</td>
<td>94</td>
<td>Me obligó a salir</td>
</tr>
<tr>
<td>AYUDAR</td>
<td>Inducción</td>
<td>70</td>
<td>Me ayudó a triunfar</td>
</tr>
<tr>
<td>LLEVAR</td>
<td>Desplazamiento</td>
<td>52</td>
<td>Me llevó a ver una casa</td>
</tr>
<tr>
<td>INVITAR</td>
<td>Inducción</td>
<td>42</td>
<td>Me invitó a ver su casa</td>
</tr>
<tr>
<td>ENSEÑAR</td>
<td>Conocimiento</td>
<td>23</td>
<td>Me enseñó a leer</td>
</tr>
<tr>
<td>IMPULSAR</td>
<td>Inducción</td>
<td>13</td>
<td>Me impulsó a escribir</td>
</tr>
<tr>
<td>ANIMAR</td>
<td>Inducción</td>
<td>10</td>
<td>Me animó a escribir</td>
</tr>
<tr>
<td>SACAR</td>
<td>Desplazamiento</td>
<td>10</td>
<td>Me sacó a bailar/lo sacó a bailar</td>
</tr>
<tr>
<td>FORZAR</td>
<td>Obligación</td>
<td>9</td>
<td>Me forzó a abandonar el intento</td>
</tr>
<tr>
<td>MANDAR</td>
<td>Desplazamiento</td>
<td>9</td>
<td>Me mandó a comprar pan</td>
</tr>
</tbody>
</table>
Paths of Schematization

Grouping verbs in classes
Two main paths of schematization / generalization:

- A verb is related, by its lexical meaning, with other partly similar verbs
  - enseñar-1 'show'
    Æ mostrar, ver, 'see', mirar 'look'
  - enseñar-2 'teach'
    Æ aprender 'learn', estudiar 'study', saber 'know'

- On the other hand, by being used in a syntactic schema, it is
  semantically construed as other verbs that realize the same pattern
  - enseñar <Subj – DO – IO>
    Æ dar 'give', decir 'say', contar 'tell', preguntar 'ask', ...
  - enseñar <Subj – Obj - a Infinitive>
    Æ dar 'give', decir 'say', contar 'tell', preguntar, ...

That puts each verb in a complex network of semantic relations

Semantic Classes of Verbs/Preds

Two main criteria of semantic classification of verbs
(both of them used in RRG)

- Aktionsart classes (based on LS)
  - State, Activity, Achievement, Accomplishment, etc

- Ontological/conceptual classes
  (based on the 'constant' part of LS)
  - Location, perception, cognition, consumption, etc.

ADESSE Verb Classes

- Goal of ADESSE verb classification:
  to represent generalizations over types of conceptual frames evoked by individual verbs

- It is a conceptual/ontological classification, inspired in lexical relations of synonymy and hyponymy/troponymy, not aspectual nor primarily syntactic

- It is a hierarchical classification, with up to four levels at the present stage
  - Top level classes 6 options
    [~Halliday’s 'process types']
  - Classes recognized so far 60 options

- With the possibility of increase granularity in the future
Semantic roles and verb classes

- Each (sub)class is associated with a set of semantic roles prototypical for the cognitive domain evoked

<table>
<thead>
<tr>
<th>Class</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>–</th>
<th>–</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feeling</td>
<td>Sensor</td>
<td>Stimulus</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perception</td>
<td>Initiator (causer)</td>
<td>Perceiver</td>
<td>Perceived</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cognition</td>
<td>Initiator (causer)</td>
<td>Cognizer</td>
<td>Content</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Possession</td>
<td>Possessor</td>
<td>Possessed</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transfer</td>
<td>Initial possessor (Donor)</td>
<td>Final possessor (Receiver)</td>
<td>Possessed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Displacement</td>
<td>Initiator (causer)</td>
<td>Theme</td>
<td>Goal</td>
<td>Source</td>
<td></td>
</tr>
<tr>
<td>Localization</td>
<td>Initiator (causer)</td>
<td>Theme</td>
<td>Locative</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Change</td>
<td>Agent</td>
<td>Patient</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Communication</td>
<td>Sayer</td>
<td>Message</td>
<td>Receiver</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Semantic roles

- There is no single list of semantic roles, and role definition is made at three levels:

  **Verb-specific roles** (representing ‘valency potential’)
  - Enseñar → A0:Teacher, A1: Learner; A2: Content

  **Class-specific roles**
  - Communication → Sayer, Message, Receiver, Topic
  - Cognition → Cognizer, Content

  **SynSem Schemas** (‘valency realizations’), pointing to verb-specific roles
  - Active Suj=A0 – DObj=A1 – IObj=A2
  - Le escríba canciones de amor

Semantic roles, verbs and semantic classes

- Each verb entry is associated with a set of arguments embracing any possible core participant with this verb
- By default, verb arguments inherit the role labels from the class(es) to which the verb belongs

<table>
<thead>
<tr>
<th>Creation</th>
<th>Creator</th>
<th>Effected</th>
<th>Receiver</th>
<th>Topic</th>
</tr>
</thead>
</table>

- But, sometimes, verb-specific role labels are also used

ADESSE roles

**CLASS-SPECIFIC ROLES**

<table>
<thead>
<tr>
<th>Experiencer</th>
<th>Perceiver</th>
<th>Cognizer</th>
<th>Stimulus</th>
<th>Possessor</th>
<th>Possessed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emoter</td>
<td>Perceived</td>
<td>Cognizer</td>
<td>Emoted</td>
<td>Perceived</td>
<td>Content</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**VERB-SPECIFIC ROLES**

- pred’(x) → BECOME pred’(x)
- pred’(x, y) → BECOME pred’(x, y)
- [do’(z)] CAUSE [BECOME pred’(x, y)]

**Increasing generalization**

- x = A0 (initiator or causer)
- y = A2 (second argument of pred')

That is only the default numbering, because many verbs have a more complex semantic structure

Common LS templates of lexical entries

- pred’(x) → BECOME pred’(x)
- pred’(x, y) → BECOME pred’(x, y)
- [do’(z)] CAUSE [BECOME pred’(x, y)]

Default indices for arguments

- z = A0 (initiator or causer)
- x = A1 (first argument of pred’)
- y = A2 (second argument of pred’)

RRG: Generalized Semantic Roles

- Verbs specific roles
- Thematic relations
- Generalized semantic roles

Increasing Generalization

- pred’(x) → BECOME pred’(x)
- pred’(x, y) → BECOME pred’(x, y)
- [do’(z)] CAUSE [BECOME pred’(x, y)]

Default indices for arguments

- z = A0 (initiator or causer)
- x = A1 (first argument of pred’)
- y = A2 (second argument of pred’)

That is only the default numbering, because many verbs have a more complex semantic structure
Generalized Semantic Roles

- \( z = A_0 \) (initiator or causer)
- \( x = A_1 \) (first argument of \( \text{pred}' \))
- \( y = A_2 \) (second argument of \( \text{pred}' \))

A0, A1, and A2 are the closest Adesse's relatives of macro-roles Actor and Undergoer.

Generalized Semantic Roles: A0 – A1 – A2 vs Actor – Undergoer

The ADESSSE hierarchy A0 – A1 – A2 is similar to the Actor-undergoer hierarchy:

- **Actor**
  - Arg of 1st arg. of \( \text{pred}' \)
  - \( \text{do}'(x,...,\text{pred}'(x,y)) \)
  - \( \text{pred}'(x,y) \)
  - \( \text{pred}'(x) \)

- **Undergoer**
  - Arg of 2nd arg of \( \text{pred}' \)
  - \( \text{caused by do}'(x,...) \)
  - \( \text{do}'(x,...,\text{pred}'(x,y)) \)
  - \( \text{pred}'(x,y) \)
  - \( \text{pred}'(x) \)

But A0-A1-A2 should not be confused with macroroles themselves.

GSR and Argument realization

- **Ditransitive**: Subj DObj IOObj (+ oblique)
  - S=0 D=2 I=1: 36%
  - S=1 D=2 I=3: 30%
  - other / not set: 34%

- **Transitive**: Subj DObj (+ oblique)
  - S=1 D=2: 61%
  - S=0 D=2: 25%
  - S=0 D=1: 3%
  - Other: 10%

- **Intransitive**: Subj (+ oblique)
  - S=1: 93%
  - S=0: 5%
  - Other: 2%

GSR and Indirect Objects

- The status of IOObj in this hierarchy is unclear.
  - In <S D I> schemas, it can be conceived as a middle (A1) or as an additional (A3) argument.
  - In <S I>, IO usually outranks the subject, as in psychological verbs [A1:Experiencer – A2:Stimulus]
    - A María le gusta la música
    - "Mary likes music"
  - Problem: the nature of IO and DO

Psychological verbs (feeling, desire, …) <A1:Experiencer – A2:Stimulus>

- a) <Subj=A1 – DDO=A2>: querer, temer, amar, odar, admirar
  - Ex: Él le quiere "He loves her"
  - \( \text{pred}(A1, A2) \); stative with human EXP as PSA
- b) <IO=A1 – Subj=A2>: gustar, interesar, importar, encantar, doler, ...
  - Ex: A ella le gusta "She likes him/her"
  - \( \text{pred}(A1, A2) \); stative with some PSA properties on IO
- c) <Subj=A2 – DO=A1>: sorprender, asustar, preocupar, impresionar, ...
  - Ex: Él la impresionó "He impressed her"
  - \( \text{pred}(A1, A2) \); stative with some PSA properties on IO

- but there are no clear limits between (b) and (c)
Psychological verbs and dative case

- Percentage of dative case for 3rd person clitic Experiencer in two-participant clauses with verbs of 'feeling'
  
<table>
<thead>
<tr>
<th>Verb</th>
<th>3rd Person Clitic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gustar</td>
<td>100% (224/225)</td>
</tr>
<tr>
<td>Moléstas</td>
<td>93% (14/15)</td>
</tr>
<tr>
<td>Impresor</td>
<td>75% (9/12)</td>
</tr>
<tr>
<td>Soprendr</td>
<td>74% (14/19)</td>
</tr>
<tr>
<td>Alegra</td>
<td>67% (2/3)</td>
</tr>
<tr>
<td>Distray</td>
<td>60% (3/5)</td>
</tr>
<tr>
<td>Preocupa</td>
<td>57% (4/7)</td>
</tr>
<tr>
<td>Tranquil</td>
<td>55% (6/11)</td>
</tr>
<tr>
<td>Consola</td>
<td>43% (3/7)</td>
</tr>
<tr>
<td>Calmar</td>
<td>17% (1/6)</td>
</tr>
</tbody>
</table>

The main conditioning factors seem to be dynamicity and effectiveness: calm, for example, is more dynamic and effective than tranquilizar.

Variable case marking

- Some 'Objects' are referred by a dative clitic [le(s)] instead of an accusative clitic [lo(s)/la(s)], even with the same verb and within the same text
  - a. El padre lo enseñaba a conocer las hierbas (Jov: 023)
  - b. A coser la enseñaban desde muy pequeña (Usos: 071)

- a. Lo que realmente lo preocupaba era… (Hist: 131)
  - b. Esos bienes que tanto le preocupan … (Hist: 70)

- Le is the canonical form for masculine and feminine 'Indirect Objects'
  - Le regalé un libro a María

- In two-participant clauses we can rank the verbs from more dative-like to less dative-like

Variable object marking

- Some Objects are doubled by a cross-reference pronominal clitic
  - ¿Conocés a Elena Garro? […]
  - ¿Y de dónde la conocés vos a Elena Garro? (BAIRES:418, 24-6)

- Doubling by clitic is usual with 'Indirect Objects'
  - Le regalé un libro a María "I gave Mary a book"
    [besides syntactic function, the main conditioning factor is accessibility status (Belloro, yesterday)]

- Some Objects are marked by preposition a
  - Encontré un amigo "I met a friend"
  - Le regalé un libro a María "I gave Mary a book"

- Preposition a is used obligatorily to mark 'full' Indirect Objects
  - Le regalé un libro a María "I gave Mary a book"
Object variation

- No dialect of Spanish has categorical rules for the use of "a", clitic doubling or "leísmo". Everywhere we have a gradient.

- The general tendency is to have unmarked nominals for O in ditransitive clauses and for P low in the animacy hierarchy. In general, we have morphologically marked objects for referents high in the animacy hierarchy (Silverstein 1976).

- Other relevant factors (not explored today):
  - "Clitic doubling" is also governed by the animacy hierarchy, but correlates more strongly with discourse status: topicality and accessibility of the referent.
  - "Case" is also governed by dialect variation, gender, and type of process (dinamicity, effectiveness).

Object and markedness

<table>
<thead>
<tr>
<th>Transitive clauses</th>
</tr>
</thead>
<tbody>
<tr>
<td>SUBJECT</td>
</tr>
<tr>
<td>Human</td>
</tr>
<tr>
<td>Highly accessible</td>
</tr>
<tr>
<td>Topic [Agent]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Objects</th>
</tr>
</thead>
<tbody>
<tr>
<td>(unmarked)</td>
</tr>
<tr>
<td>Non Human</td>
</tr>
<tr>
<td>Indefinite</td>
</tr>
<tr>
<td>Low accessibility</td>
</tr>
<tr>
<td>(Part of) Rheme</td>
</tr>
<tr>
<td>Patient</td>
</tr>
</tbody>
</table>

Conclusions

ADESSE: What is it good for?

- ADESSE is, above all, a database for the empirical study of the interaction between verbs and constructions in Spanish:
  - Constructional alternatives for a verb, a syntactic function or a semantic role (with frequencies in a corpus).
  - Verbs and syntactic constructions for a semantic domain.
  - Verbs and semantic domains for a particular construction.
  - ... Additionally, it allows the search and study of many imaginable correlations between syntactic and semantic features (case, person, number, definiteness, tense, mood, ...)
Conclusions / final remarks

• It is necessary to observe (spoken and written) language in use and to take into account frequency as an important factor of language structure and meaning.

• Many grammatical categories manifest as a gradient and not as discrete categories. We have seen that this is the case with argument structure, argument realization, and grammatical relations. The task is to identify the factors influencing the choice of a form, and the strength of each factor.

Conclusions / final remarks

• It order to achieve descriptive adequacy, we need corpora annotated with increasingly detailed syntactic and semantic categories.

• But the categories used in corpus annotation should not be taken for granted, and must be revised in the light of corpus evidence.

• Therefore, we have to move continually from analytical categories to corpus and from corpus to analytical categories.

• (That is also a disclaimer: Researchers can get adesee data at <http://adesee.uvigo.es/data/> with some analysis, but the final analysis of each example is responsibility of the user).

References


• Bybee, Joan. (2006) From Usage to Grammar The Mind’s Response to Language. CUP.


