

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

García-Miguel: Corpus-based approach to argument structure

RRG 2007

2

Functional Grammar(s)

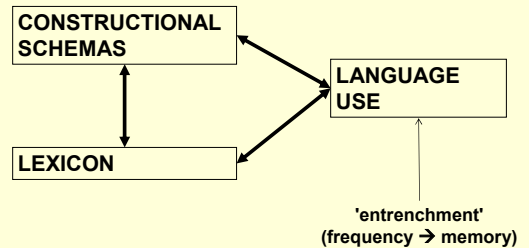
- This talk is not specifically about RRG, but it takes as background many ideas shared by most functionalists
- Functionalists share the idea that grammar associates forms with meanings and discourse functions:
 - language is a "system of communicative social action in which grammatical structures are employed to express meaning in context" (Van Valin 2005: 1)
- They differ with respect to (among other things):
 - Standards of adequacy (typological, psychological, ...)
 - How they conceive the relation *system - use*
 - The role of formalization and the specifics of the formalismModerate functionalists (RRG) ↔ Extreme functionalists

García-Miguel: Corpus-based approach to argument structure

RRG 2007

3

(A simplified model of) Functional grammar



García-Miguel: Corpus-based approach to argument structure

RRG 2007

4

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)

García-Miguel: Corpus-based approach to argument structure

RRG 2007

5

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

García-Miguel: Corpus-based approach to argument structure

RRG 2007

6

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 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

[USC 1990-1999]

[Univ. of Vigo 2002-]

Syntactic information

Grammatical features of clauses, verbs and arguments of the corpus

[Each record (clause): 64 fields]

All the syntactic information from BDS

+
Semantic information

- Verb senses
- Verb classes
- Semantic roles

159.000 clauses

3.500 verbs

13.500 valency patterns

ADESSE

REGALAR

Acep: 1.0:Obsequiar. Dar algo de forma gratuita.

Locucional:

Clasificación: [222] Transferencia

Figurado:

Activa

S D I

San Din lan

ORDEN IVC

ARGUMENTOS

RASGOS GENERALES

F	L	Rol	FUNCIÓN	CONC	PREP	NUCLEO	UNIDAD
<input type="checkbox"/>	0-	[POSEEDOR-INICIAL]	S Sujeto (implicito)	3 ^a pl			
<input type="checkbox"/>	2-	[POSESIÓN]	D Obj Directo				FN
<input type="checkbox"/>	1-	[POSEEDOR-FINAL]	I Obj Indirecto	le (masc.)			FN

para recibir a los recién casados. Al novio le regalaron un automóvil convertible con su nombre grabado en letras góticas bajo el escudo de la fábrica.

Más ... CRONICA: 044, 01

(part of) a record in ADESSE

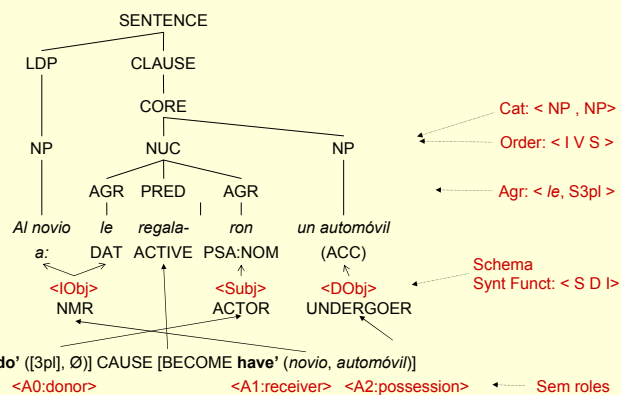
Field	Value		
TEXT	<i>Al novio le regalaron un automóvil convertible</i> [CRO: 44, 1] 'They gave the bridegroom a convertible car as gift'		
PRED	REGALAR		
Verb class	Transfer of possession		
Voice	Active		
	A0	A1	A2
Synt Function	Subj	DObj	IObj
Agreement	3pl		le
Synt Category		NP	NP
Sem Role	Donor	Possession	Receiver
		<i>automóvil</i>	<i>novio</i>
Synt Schema	S D I		
Order	IVD		

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 selects a subset of the potential participants of the scene evoked by the verb

- Juan* [0] *le escribió una carta* [1] *a su madre* [2] *sobre sus recuerdos de infancia* [3]
'John wrote a letter to his mother about his childhood remembrances'
- Juan* [0] *escribió una carta* [1]
'John wrote a letter'
- 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'

- a) *Deco* [1] *sustituyó a Xavi* [2]
'Deco replaced Xavi'
- b) *Rijkaard* [0] *sustituyó a Xavi* [2]
'Rijkaard replaced Xavi'
- c) *Rijkaard* [0] *sustituyó a Xavi* [2] *por Deco* [1]
'Rijkaard replaced Xavi with Deco'

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

– **Enseñar** 'teach'

- a) *Ella* [0] *le* [1] *enseñaba su idioma* [2]
'She taught him her language'
- b) *Ella* [0] *enseñó al niño* [1] *a caminar* [2]
'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)

ESCRIBIR 1 *Poner un mensaje por escrito*

Clasificación

0	Escritor[AGENTE]/(Emisor)
1	Texto[EFFECTUADO]/(Mensaje)
2	Receptor[BENEFICIARIO]/(Receptor)
3	Asunto/(Asunto)

1 cambio: creación
2 comunicación: general

VOZ	Esquema	S	D	I	O	O2	A	P	Ejs
Activa	S	0							58
Activa	S D	0	1						141
Activa	S D I	0	1	2					21
Activa	S I	0		2					58
Activa	S R(de)	0			3				4
Activa	S R(sobre)	0			3				4
Pasiva	S	1							3
Pasiva	S A	1					0		4
Pasiva	S I A	1	2				0		1

Registro: 1800 de 3755

Valency patterns and frequency

Valency potential of the verb ENSEÑAR 'teach'

A0 (Teacher)	A1 (Learner)	A2 (Content)	N	Examples
Valency patterns in Active Voice				
A0:Subj	A1:IObj	A2:DObj	58	<i>Me enseñaba su idioma</i> [JOV:134]
A0:Subj	A1:IObj		28	<i>Si su hijo no sabe, él le enseñará</i> [SON:99]
A0:Subj	A1:IObj	A2:Obl(a)	22	<i>Eso me enseñará a fiarme de ti</i> [PAI:161]
A0:Subj		A2:DObj	18	<i>Enseñaban cosas útiles</i> [MAD:277]
A0:Subj			4	<i>Kant enseñó en Königsberg</i> [TIE:195]
A0:Subj		A2:Obl(a)	2	<i>Enseñaban a saber comer</i> [MAD:277]
A0:Subj	A1:DObj	A2:Obl(a)	1	<i>A coser la enseñaban desde pequeña</i> [USOS:71]

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 and gradience

Arguments of *Enseñar* 'teach' (N clauses =139)

Argument	Count	Percentage
0 Teacher	138	(99.3 %)
1 Learner	111	(79.9 %)
2 Content	105	(75.5 %)

Arguments of *Escribir* 'write' (N clauses = 321)

Argument	Count	Percentage
0 Writer	300	(93.5 %)
1 Text	208	(64.8 %)
2 Receiver	85	(26.5 %)
4 Topic	10	(3.1 %)

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 accomodation. 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
- Nevertheless, this is a 'false dichotomy' (Croft 2003), motivated either by the level of granularity or by the perspective adopted when linking syntax and semantics (Van Valin 2004)

Lexical entries in ADESSE

Two levels:

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

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

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)

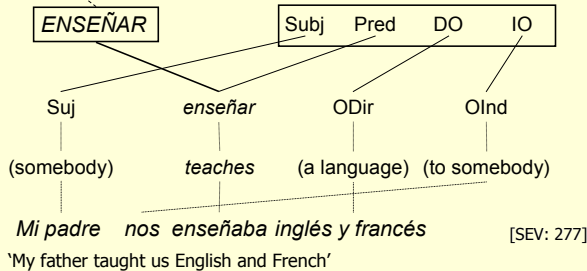
Unifying verb senses

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

teaching verbs/frame



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

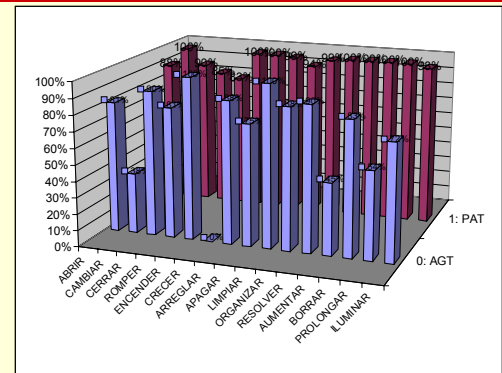
- Many lexicalist approaches have focused on whether verbs admits a certain alternation, for ex. the causative alternation
- (or, alternatively, whether they can be combined with a certain type of argument)

	A1	A0 – A1
<i>Cambiar</i> 'change'	Yes	Yes
<i>Crecer</i> 'grow'	Yes	No

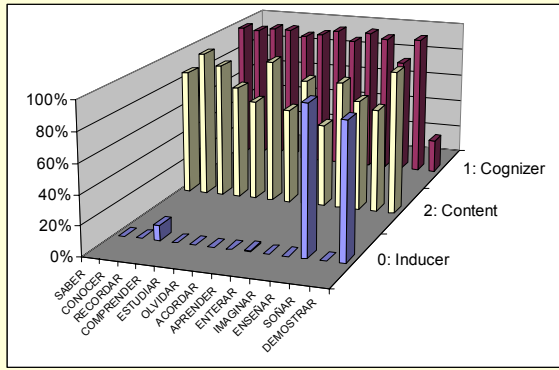
	A0 – N1 – N2	N1 – N2
<i>Aprender</i> 'learn'	No	Yes
<i>Enseñar</i> 'teach'	Yes	Yes

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



Arguments and behavior profile: Verbs of knowledge



Valency alternations and behavior profile

- Each lexical element can be described in terms of the features and constructions it combines with in context (its 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
- It is hypothesized that the differences in frequency are motivated by differences in meaning

"Surface generalizations"

» (cf. Goldberg 1995, 2002; also Dowty 2000)

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 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

Verbs in the schema <Subj DO IO>

VERB	Meaning	N	Example
DAR	'Give'	1272	<i>Mi padre sólo me da dinero para estudiar</i> [SON:115]
DECIR	'Say, tell'	599	<i>Le dije que no quería volver</i> [TER: 046]
HACER	'Make'	514	<i>Le hizo la promesa de llevarte al local</i> [TER: 119]
CONTAR	'Tell'	308	<i>Rosetta le contaba que el otro iba empeorando</i> [SON: 233]
PEDIR	'Ask, request'	273	<i>Nos pidió que lo esperáramos</i> [HIS: 013]
PREGUNTAR	'Ask, inquire'	219	<i>Me preguntó que si tenía dinero</i> [LAB: 268]
PERMITIR	'Allow'	124	<i>¿Me permite usar su teléfono?</i> [SON: 285]
OFRECER	'Offer'	119	<i>No puedo ofrecerles grandes comodidades</i> [LAB: 115]
EXPLICAR	'Explain'	112	<i>Le explicó cómo funcionaba la imprenta</i> [TER: 062]
PONER	'Put'	101	<i>Si se te cae un ojo, te pondrán otro enseguida</i> [MOR: 094]
TRAER	'Bring'	96	<i>Le trae un kilo de bombones a mamá</i> [BAI:424]
DEJAR	'Lend'	96	<i>El profesor le dejaba la casa para que la habitara</i> [HIS:169]

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

	CLASS	Clauses	Verbs	Example verbs
22	POSSESSION	2568	80	<i>dar</i> 'give'
4	COMMUNICATION	2419	108	<i>decir</i> 'say, tell'
6	MODULATION	806	18	<i>hacer</i> 'do, make'
31	LOC & MOVEMENT	686	88	<i>traer</i> 'bring'
1	MENTAL	615	64	<i>recordar</i> 'remember'
32	CHANGES	430	140	<i>abrir</i> 'open'
33	OTHER FACTS	305	80	<i>tocar</i> 'touch'
21	ATTRIBUTIVE	180	17	<i>costar</i> 'cost'
5	EXISTENTIAL	118	20	<i>(causar</i> 'cause')
		8127	615	

S D/lan a Inf

More frequent verbs in the constructional schema **S D/lan a Inf** (ADESSE):

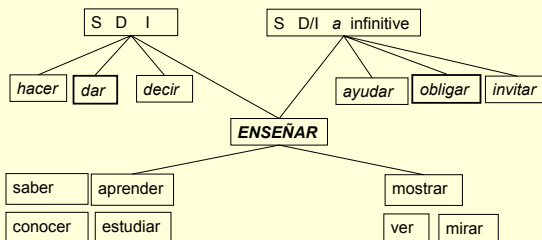
VERB	Class	N	Example
OBLIGAR	Obligación	94	<i>Me obligó a salir</i>
AYUDAR	Inducción	70	<i>Me ayudó a triunfar</i>
LLEVAR	Desplazamiento	52	<i>Me llevó a ver una casa</i>
INVITAR	Inducción	42	<i>Me invitó a ver su casa</i>
ENSEÑAR	Conocimiento	23	<i>Me enseñó a coser</i>
IMPULSAR	Inducción	11	<i>(esto) Me impulsó a escribir</i>
ANIMAR	Inducción	10	<i>Me animó a escribir</i>
SACAR	Desplazamiento	10	<i>Me sacó a bailar//lo sacó a saludar</i>
FORZAR	Obligación	9	<i>Me forzó a abandonar el intento</i>
MANDAR	Desplazamiento	9	<i>Me mandó a comprar pan</i>

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'
 - *mostrar* 'show', *ver* 'see', *mirar* 'look'
 - enseñar-2* 'teach'
 - *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>
 - *dar* 'give', *decir* 'say', *contar* 'tell', *preguntar* 'ask', ...
 - enseñar* <Subj – Obj – a Infinitive>
 - *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 [~Halliday's 'process types'] 6 options
 - Classes recognized so far 60 options
- With the possibility of increase granularity in the future

ADESSE hierarchy of semantic classes

		Examples	
1 MENTAL	11 Feeling	<i>gustar</i> , <i>temer</i>	
	12 Perception	<i>ver</i> , <i>escuchar</i>	
	13 Cognition	<i>saber</i> , <i>enseñar</i>	
	131 Knowledge	<i>creer</i>	
	132 Belief	<i>creer</i>	
2 RELATIONAL	21 Attributive	<i>ser</i>	
	22 Possession	<i>tener</i> , <i>dar</i>	
3 'MATERIAL'	31 "Space"	311 Displacement	<i>ir</i> , <i>llevar</i>
		312 Location	<i>estar</i> , <i>poner</i>
	32 Change	321 Creation	<i>hacer</i> , <i>crear</i>
		322 Modification	<i>romper</i> , <i>cambiar</i>
		323 Destruction	<i>destruir</i>
		33 Other facts	<i>tocar</i>
	4 COMMUNICATION	41 Judgement	<i>decir</i> , <i>hablar</i>
42 Command		<i>criticar</i>	
5 EXISTENTIAL		<i>haber</i> , <i>aparecer</i>	
6 MODULATION	60 Causative	<i>hacer</i> , <i>obligar</i>	
	61 Dispositive	<i>tratar</i> , <i>atreverse</i>	

ADESSE top-level classes

		Exs	VERBS	CLAUSES
1 MENTAL	11 Feeling	<i>Gustar</i> 'like'	271	7256
	12 Perception	<i>Ver</i> 'see'	115	13148
	13 Cognition	<i>Saber</i> 'know'	171	15445
2 RELATIONAL	21 Attributive	<i>Ser</i> 'be'	232	17615
	22 Possession	<i>Tener</i> 'have'	185	13149
3 'MATERIAL'	31 "Space"	<i>Ir</i> 'go'	629	27641
		<i>Abrir</i> 'open'	923	10964
	33 Other facts	<i>Tocar</i> 'touch'	429	905
	34 Behaviour	<i>Reír</i> 'laugh'	331	3960
4 COMMUNICATION		<i>Decir</i> 'say'	335	15521
5 EXISTENTIAL		<i>Haber</i> 'exist'	201	11777
6 MODULATION		<i>Hacer</i> 'make'	139	11336
TOTAL			3978	158361

Semantic roles and verb classes

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

Class	0	1	2	--	--
Feeling		Senser	Stimulus		
Perception	Initiator (causer)	Perceiver	Perceived		
Cognition	Initiator (causer)	Cognizer	Content		
Possession		Possessor	Possessed		
Transfer	Initial-possessor (Donor)	Final possessor (Receiver)	Possessed		
Displacement	Initiator (causer)	Theme		Goal	Source
Localization	Initiator (causer)	Theme		Locative	
Change	Agent	Patient			
Communication		Sayer	Message	Receiver	

Semantic roles

- There is no single list of semantic roles, and role definition is made at three levels:
 - Verb-specific roles** (representing 'valency potential')
 - Escribir* → A0:Writer, A1:Text, A2:Receiver, A3:Topic
 - 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 escribía canciones de amor

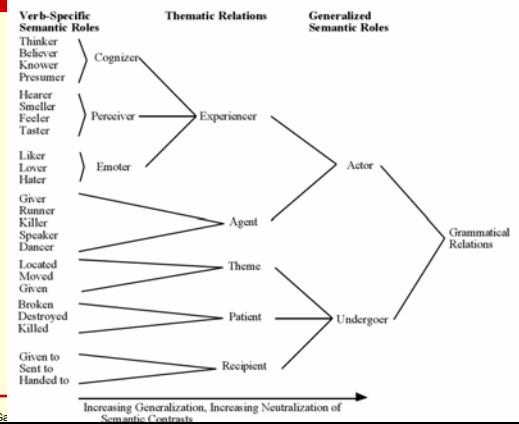
Semantic roles, verbs and semantic classes

- 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

CREATION	Creator	Effected		
COMMUNICATION	Sayer	Message	Receiver	Topic
<i>Escribir</i> 'write'	A0	A1	A2	A3
	Writer	Text		

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

RRG: Generalized Semantic Roles



ADESSE roles

CLASS-SPECIFIC ROLES	VERB-SPECIFIC ROLES
Experiencer	Emoter — <i>gustar.01, temer.01</i>
	Perceiver — <i>ver.01, mirar.01</i>
	Cognizer — <i>saber.01, enseñar.01, crear.01</i>
Stimulus	Emoted — <i>gustar.02, temer.02</i>
	Perceived — <i>ver.02, mirar.02</i>
	Content — <i>saber.02, enseñar.02, crear.02</i>
Possessor — <i>tener.01, dar.01</i>	
Possessed — <i>tener.02, dar.02</i>	
Patient	Created — <i>hacer.01, crear.01</i>
	Affected — <i>romper.01, cambiar.01</i>
	Destroyed — <i>destruir.01</i>

← Increasing generalization

Generalized Semantic Roles (still a tentative in ADESSE)

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'**)

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

Generalized Semantic Roles

- z = A0 (initiator or causer)
- x = A1 (first argument of **pred'**)
- y = A2 (second argument of **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 ADESSE hierarchy A0 – A1 – A2 is similar to the Actor-undergoer hierarchy:

	ACTOR		UNDERGOER
Arg of	1 st arg. of	1 st arg. of	2 nd arg. of
DO	do' (x, ...	pred' (x,y)	pred' (x,y) pred' (x)
A0		A1	A2
Arg of	1 st arg. of	1 st arg. of	2 nd arg. of
[do' (x,ø)]	CAUSE ...	do' (x, ...	pred' (x,y) pred' (x,y) or pred' (x)

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

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

•RRG: Default macroroles for *saber*, *aprender* and *enseñar*:

	Actor:	Undergoer:	NMR:
<i>SABER</i> < know' (x,y)>	<i>Knower</i>	<i>Thing known</i>	
<i>APRENDER</i> <INGR know' (x,y)>	<i>Learner</i>	<i>Thing learned</i>	
<i>ENSEÑAR</i> <[do' (x, ø)] CAUSE [BECOME know' (y,z)]>	<i>Teacher</i>	<i>Thing learned / Learner</i>	<i>Learner / Thing learned</i>

•ADESSE: Arguments of *saber*, *aprender* and *enseñar*:

	A0	A1	A2
<i>SABER</i> < know' (x,y)>		<i>Knower</i>	<i>Thing known</i>
<i>APRENDER</i> <INGR know' (x,y)>		<i>Learner</i>	<i>Thing learned</i>
<i>ENSEÑAR</i> <[do' (x, ø)] CAUSE [BECOME know' (x,y)]>	<i>Teacher</i>	<i>Learner</i>	<i>Thing learned</i>

GSR and Argument realization

Ditransitive: Subj DObj IObj (+ oblique)

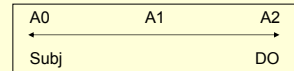
S=0 D=2 I=1	36%
S=1 D=2 I=3	30%
other / not set	34%

Frequency of argument realizations in **Active voice**

• Subject is almost always higher than DO in the hierarchy of GSRs

Transitive: Subj DObj (+ oblique)

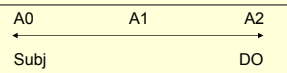
S=1 D=2	61%
S=0 D=1	25 %
S=0 D=2	3 %
Other	10%



Intransitive: Subj (+ oblique)

S=1	93%
S=0	5%
Other	2%

GSR and Indirect Objects



• The status of IObj 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

Three participants:

Subj DObj IObj (+ oblique)	
S=0 D=2 I=1	36%
S=1 D=2 I=3	30%
other / not set	34%

Two participants:

Subj IObj (+ oblique)	
S=2 I=1	61%
S=1 I≥2	25%
Other	10%

GSR and Argument realization

"Inversion (?)"

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

a) <Subj=A1 – DO=A2>: *querer, temer, amar, odiar, admirar*,

Ex: *Él la quiere* "He loves her"

– **pred'**(A1, A2); stative with human EXP as PSA

b) <IO=A1 – Subj=A2>: *gustar, interesar, importar, encantar, doler*, ..

Ex: *Ella le gusta* "She likes him/her"

– **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"

– → <S=A0:Causer D=A1:Experiencer>

• 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'

A Paula le/la alegró la noticia
 'The news pleased Paula'

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

<i>Gustar</i> 'like'	100%	(224/225)
<i>Molestar</i> 'bother'	93%	(14/15)
<i>Impresionar</i> 'impress'	75%	(9/12)
<i>Sorprender</i> 'surprise'	74%	(14/19)
<i>Alegrar</i> 'please'	67%	(2/3)
<i>Distraer</i> 'amuse'	60%	(3/5)
<i>Preocupar</i> 'worry'	57%	(4/7)
<i>Tranquilizar</i> 'calm'	55%	(6/11)
<i>Consolar</i> 'console'	43%	(3/7)
<i>Calmar</i> 'calm'	17%	(1/6)

Objects

- Subj & DO & IO = direct core arguments
- Rough RRG equivalents
 - Subject = PSA
 - DO = Undergoer
 - IO = NMR
- Corpus data reveal a continuum DO – IO with several factors at play:
 - Case: accusative vs dative (3rd person clitic)
 - Preposition *a*
 - Cliticization and clitic-doubling

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 le enseñaba a conocer las hierbas* (Jov: 023)
 The father taught him [dat] to know about herbs
 - b. *A coser la enseñaban desde muy pequeña* (Usos: 071)
 She was taught to sew since her early infancy
 - a. *Lo que realmente lo preocupaba era...* (Hist: 131)
 What actually worried him [acc] was ...
 - b. *Esos bienes que tanto le preocupan...* (Hist: 70)
 These goods that so much worry him [dat]
- Le* is the canonical form for masculine and feminine 'Indirect Objects' [R] in ditransitive clauses
 - 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)]

Variable object marking

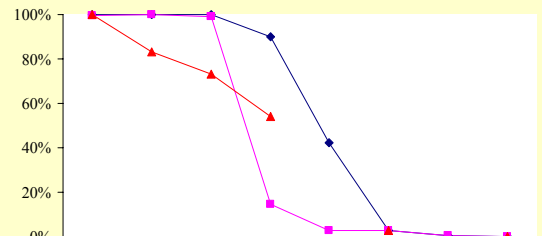
- Some Objects are marked by **preposition a**

Encontré a un amigo "I met a friend"
Encontré un amigo id.

- Preposition *a* is used obligatorily to mark 'full' Indirect Objects

Le regalé un libro a María
 'I gave Mary a book'

	2-participants: Subj – Obj	3-participants Subj – DO – IO	
	Object	O = DO	R = IO
A + NP [vs NP]	9.9 %	0.7%	100 %
Clitic Doubling [vs (a) NP]	3.8%	1.0%	42.8%
Dative clitic [vs Accusative]	31.1%	0.3%	99.7%



	1ª 2ª	Vd	Pro 3ª	anim def	anim indef	inan def	inan indef	Clause
◆ % a	100%	100%	100%	90%	42,4%	2,7%	0,4%	0%
■ % doubling	99,5%	100%	99,2%	14,7%	2,8%	2,7%	0,3%	0,2%
▲ % dative	100%	83%	73%	54%		2,7%		0,2%

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).

Subject and object in Spanish

Participants in SUBJ-PRED-OBJ [+Obl] pattern [N = 81954 clauses]

	SUBJ	OBJ
Human	80.50 %	27.14 %
Agreement/clitic only	63.80 %	25.90 %
Definite (if NP)	90.00 %	66.33 %
Preverbal (if NP or clause)	73.67%	3.84 %

Participants in SUBJ-PRED-DO-IO (+Obl) pattern [N = 8455 clauses]

	SUBJ	DO	IO
Human	84.18 %	2.25%	90.24 %
Agreement/clitic only	65.95 %	10.65%	74.14 %
Definite (if NP)	90.06 %	53.57%	89.02 %
Preverbal (if NP or clause)	74.50 %	2.40 %	9.50 %

Object and markedness

Transitive clauses

SUBJECT	DIROBJ
Human	Non Human
Highly accessible	Low accessibility
Definite	less definite
Topic	(Part of) Rheme
[Agent]	[Patient]

Objects

(unmarked)	(marked)	
Non Human	Human	→ a-marking, /e
Indefinite	Definite	
Low accessibility	More accessible	→ 'doubling'
(Part of) Rheme	Topic	
Patient	Less affected	→ 'leísmo'

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

- In 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 adesse data at <http://adesse.uvigo.es/data/> with some analysis, but the final analysis of each example is responsibility of the user)

References

- Belloro, Valeria (yesterday). The pragmatics of dative doubling in Spanish. 2007 RRG Conference. México Butler, Christopher S. 2004. Corpus studies and functional linguistic theories. *Functions of Language* 11/2: 147-86.
- Butler, Christopher S. (2004) Corpus studies and functional linguistic theories. *Functions of Language* 11/2: 147-86.
- Bybee, Joan. (2006) From Usage to Grammar The Mind's Response to Repetition. *Language* 82/4 711-33.
- Bybee, Joan, and Paul Hopper (2001). *Frequency and the emergence of linguistic structure*. Amsterdam: John Benjamins.
- Company, Concepción. (2001). Multiple dative-marking grammaticalization. Spanish as a special kind of primary object
- Croft, William. (2003). Lexical rules vs. constructions: a false dichotomy. *Motivation in Language. Studies in honor of Günter Radden*. eds H. Cuyckens et al, 49-68. Amsterdam: John Benjamins.

- Dowty, David. (20009. 'The garden swarms with bees' and the fallacy of 'argument alternation'. *Polysemy: theoretical and computational approaches*. eds Y. Ravin, and C. Leacock, 110-128. Oxford: Oxford University Press.
- Fillmore, Charles J. (1992). "Corpus linguistics" or "Computer-aided armchair linguistics". *Directions in Corpus Linguistics*. ed Jan Svartvik, 35-60. Berlin: Mouton de Gruyter.
- Goldberg, Adele E. (1995). *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago / London: University of Chicago Press.
- Goldberg, Adele E. (2002). Surface generalizations: An alternative to alternations. *Cognitive Linguistics* 13/4: 327-56.
- Gries, Stefan Th. (2003). Towards a corpus-based identification of prototypical instances of constructions. *Annual Review of Cognitive Linguistics* 1: 1-27.

- Halliday, M. A. K. (2004). *An introduction to functional grammar* (3rd edition, revised by Christian M. I. M. Matthiessen). London: Arnold.
- Kilgarriff, Adam. (1997). "I don't believe in word senses". *Computers and the Humanities* 31: 91-113.
- Langacker, Ronald W. (1991). *Foundations of Cognitive Grammar, Vol. II: Descriptive Application*. Stanford: Stanford Univ. Press.
- Levin, Beth. (1993). *English Verb Classes and Alternations: a Preliminary Investigation*. Chicago: University of Chicago Press.
- Manning, Christopher. (2003). Probabilistic Syntax. *Probabilistic Linguistics*. eds Rens Bod, Jennifer Hay, and Stefanie Jannedy, 289-341. Cambridge (Mass.): MIT Press.
- Nichols, Johanna. (1984). Functional theories of grammar. *Annual Review of Anthropology* 13: 97-117.
- Silverstein, Michael. (1976). Hierarchies of features and ergativity. *Grammatical categories in Australian languages*. ed. R. M. W. Dixon, 112-71. Canberra Australian Institute of Aboriginal Studies.

- Thompson, Sandra A. and Paul Hopper. (2001). Transitivity, clause structure, and argument structure: evidence from conversation. *Frequency and the emergence of linguistic structure*. eds Joan Bybee, and Paul Hopper, 27-60. Amsterdam: John Benjamins.
- Van Valin, Robert D. (2001). Functional Linguistics. *The Handbook of Linguistics*. eds M. Aronoff & J. Rees-Miller, 319-36. Oxford: Blackwell.
- Van Valin, Robert D. (2005). *Exploring the Syntax-Semantics Interface*. Cambridge: Cambridge University Press.
- Van Valin, Robert D. (2004). Lexical representation, co-composition, and linking syntax and semantics. (Ms) http://linguistics.buffalo.edu/research/rrg/vanvalin_papers/LexRepCoCompLnkgRRG.pdf.
- Van Valin, Robert D., and Randy J. LaPolla. (1997). *Syntax, Structure, meaning and function*. Cambridge: Cambridge University Press.