EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

PHILOSOPHISCHE
FAKULTÄT

Semantic transparency and variation in nominal syntagmatic compounds in Romance languages

Dr. Inga Hennecke **inga.hennecke@uni-tuebingen.de**

Compounds or syntagms?

vent d'ouest – vent du nord
accident de parcours – accident de la route (Rackow 1994: 1f.)

juguete para niños
juguete de niños

un chapeau à plumes
un grand chapeau à plumes blanches (Bosredon, Tamba 1991:42)

Syntagmatic compounds - Terminology

phrasal compounds (Bisetto 2009)
syntactic compounds (Rio-Torto & Ribeiro 2009)
improper compounds (Kornfeld 2009)
phrasal lexemes (Masini 2007,2009,2012)
frozen multiword units (Guevara 2012)
lexicalized syntactic constructions (Villoing 2012)
lexicalized phrases (Fraudin 2009)
syntactic words (DiSciullo & Williams 1987)
prepositional compounds

Syntagmatic compounds - Definition

Syntagmatic compounds include generally (partially) lexicalized constructions, such as:

[N PREP N]N	*dulce de leche*, ‚caramel‘
[N PREP Art N]N	*árbol de la cera* ‚wax myrtle‘
[N + A]N	*hierbabuena* ‚mint‘
[A + N]N	*malasombra* ‚evil person’

Moyna 2011: 38

→These constructions possess an internal structure that – at first glance - appears syntactic

→These constructions also show certain characteristics of lexicalization, fixedness as well as semantic specification

Syntagmatic Compounds - Definition

Table 1. Phrasal nouns in Romance languages

Language	Types	Phrasal lexemes	Lit.	Glosses
French	[ADJ N]$_N$	*première violon*	first violin	'first violin'
	[N *à* V_{INF}]$_N$	*salle à manger*	room at eating	'dining room'
Italian	[N *da* N]$_N$	*camera da letto*	room from bed	'bedroom'
	[N ADJ]$_N$	*disco rigido*	disk rigid	'hard disk'
Portuguese	[N *de* N]$_N$	*cadeira de rodas*	chair of wheels	'wheel chair'
	[N ADJ]$_N$	*estado civil*	status civil	'marital status'
Spanish	[N ADJ]$_N$	*luna nueva*	moon new	'new moon'
	[N *de* N]	*coche de carreras*	car of races	'race/racing car'

N Prep N constructions in Romance languages may be formed by means of different internal prepositions

Table 3. N PREP N constructions in Italian

Structure	Examples	Lit.	Glosses
[N *di* N]$_N$	*casa di cura*	home of treatment	'nursing home'
[N *a* N]$_N$	*camera a gas*	room at gas	'gas chamber'
[N *da* N]$_N$	*ferro da stiro*	iron from iron	'iron'
[N *in* N]$_N$	*ballo in maschera*	ball in mask	'masked ball'
[N *per* N]$_N$	*cibo per cani*	food for dogs	'dog food'
[N *con* N]$_N$	*corsa con handicap*	race with handicap	'handicap race'
[N *su* N]$_N$	*hockey su prato*	hockey on meadow	'field hockey'

Masini 2009: 257-259

Syntagmatic compounds – word formation or syntax?

"they [e.g. *fin de semana, sabelotodo*] are clearly not formed by any rule of the language, they are **"frozen" multiword units** arising as the result of processes of lexicalization and fossilization and **do not belong in the core of word-formation**" (Guevara 2012: 179)

"a sequence of words is a **morphological compound** if it is a lexeme (noun, verb or adjective) constructed from other lexemes, according to a **non-syntactic mode of organization**" (Villoing 2012:35)

„a **very representative pattern** of phrasal compounds in Portuguese" (Rio-Torto & Ribeiro 2011)

„ Formations of this kind (…) constitute **very productive lexical patterns**" (Rainer 2016: 2624)

Syntagmatic compounds – word formation or syntax?

	Locuciones	Compuestos sintagmáticos	Colocaciones
Fijación	+	+	+/-
	oveja negra	*moscas blancas*	*batalla encarnizada*
Idiomaticidad	+	+/-	+
	lágrimas de cocodrilo	*ojo de buey* *cuello de cisne*	*odio mortal*
Lexicalización	+	+	-
	pájaro de mal agüero	*paso de cebra*	*error garrafal*
Unidad de denominación	-	+	-
	edad del pavo	*escalera de caracol*	*envidia sana*
Motivación pragmática	+	+/-	-
	olla de grillos	*diente de león*	*éxito fulgurante*
Comportamiento paradigmático	-	+	+
	boca del lobo	*oso pardo, oso panda, oso polar*, etc.	*lluvia torrencial*

Buenafuentes de la Mata 2006: 14

other delimitation tests, e.g. Bouvier 2000, Bisetto & Scalise 2005, Lieber & Scalise 2007, Masini 2009, Masini & Scalise 2012)

The role of the prepositional element

The common assumption is that the internal preposition does lack any kind of semantic content and fulfils pure linking functions:

« (…) nous aimerions préciser que **la préposition *de* ne code pas de sens propre**. Par contre, on peut l'interpréter comme **une instruction spécifique de mise en relation**. » (Bartning 1993: 164)

: « Alors **que les prépositions incolores ont pour fonction dominante de vectoriser une relation qu'elles ne codent pas » (**Cadiot 1989)

« cette préposition a pour vocation première d'exprimer l'opération de construction et de nomination d'une sous-classe (hyponyme) de N1, et non d'une simple occurrence. Du coup, elle est par nécessité associée peu ou prou a une **perte d information**, ou mieux à ce qu'on appellera un **calibrage référentiel.** » (Cadiot 1993: 196)

« les mots formés à l'aide de *à* et *de* sont aussi des mots complexes sous-classifiants, et les prépositions des « **opérateurs de couplage** ». (Bosredon & Tamba 1991 : 44),

Theoretic background – Construction grammar

(23) $[\text{N PREP Y}]_N$ < a kind of N that has to do with Y >

I_I

ABSTRACT LEXICAL CONSTRUCTION

$[\text{N1 PREP N2}]_N$ < a kind of N1 that has to do with N2 >

I_I

INTERMEDIATE LEXICAL CONSTRUCTION 1

$[\text{N1 } \textit{di} \text{ N2}]_N$ < a kind of N1 that has to do with N2 >

I_I

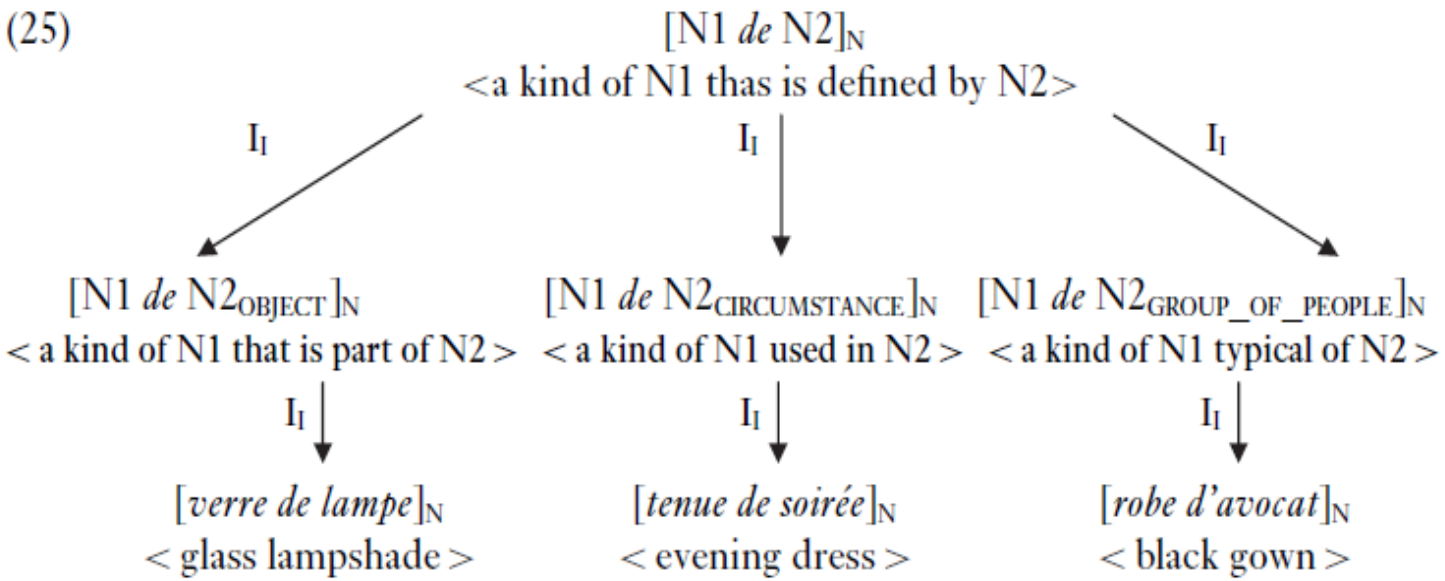
INTERMEDIATE LEXICAL CONSTRUCTION 2

$[\text{casa } \textit{di} \text{ cura}]_N$ < a kind of house with treatment >

SPECIFIED LEXICAL CONSTRUCTION (= PHRASAL LEXEME)

Masini 2009: 262 following Goldberg 1995

Theoretic background – Construction grammar



Benveniste 1966, reinterpretated by Masini 2009: 263

Syntagmatic compounds –prepositional variation

1. Variation de – para/pour

a. Sp. *fuente de horno - fuente para horno* (esTenTen)
b. Pt. *água de lavagem – água para lavagem* (ptTenTen)
c. Fr. *jouet d'enfant - jouet pour enfants* (frTenTen)

2. Variation de – a/à

a. Sp. *motores de gasolina – motores a gasolina* (esTenTen)
b. Fr. *épingle de nourrice - épingle à nourrice*
c. Pt. *Fogão de lenha - Fogão a lenha* (ptTenTen)

3. Variation de – en/em

a. Fr. *chemise de coton – chemise en coton* (frTenTen)
b. Pt. *bracelete de aço – bracelete em aço* (ptTenTen)
c. Sp. *ciclismo de pisto – ciclismo en pisto* (esTenTen)

4. Elision [N Prep N] → [N N]

a. Pt. *efeito de estufa - efeito estufa* (Faria 2010)
b. Sp. *ducha de teléfono – ducha teléfono* (Moyna 2011)
c. Fr. *stylo à bille - stylo-bille* (Fradin 2009)

Dataset – the TenTen Corpus Family

	frTenTen	esTenTen	ptTenTen
Tokens	11,444,973,582	10,994,616,207	4,626,584,246
Words	9,889,689,889	9,497,402,122	3,900,501,097
Sentences	456,065,104	407,205,587	190,221,913
Paragraphs	188,079,362	213,364,685	91,248,976
Documents	20,400,411	22,287,566	10,216,060

https://the.sketchengine.co.uk (Corpus Info)

Dataset for the analysis:

- Sample of 100 Million tokens per language (frequencies to 100 millionen tokens normalized and trimmed with a relative frequency smaller than 1 – to avoid sampling errors)

- Automatic extraction of all N Prep N constructions:

Language	Types	Tokens
French	284.432	1.301.850
Spanish	385.162	1.949.941
Portuguese	642.022	3.204.462

Type-Token-Frequenzen Datenset1 (Rohdaten)

Dataset – the TenTen Corpus Family

Datenset2:

- extraction of all paradigmatic sets (same N1 and N2 but different prepositional element)

Language	Types	Tokens
French	18.528	148.248
Spanish	15.219	366.284
Portuguese	173.370	1.477.230

Type-Token-Frequencies Datenset2

Datenset3:

- manual sorted Dataset 2:

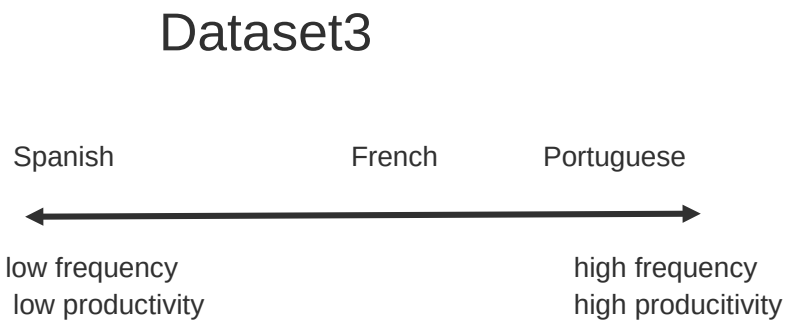
- grammaticalised construction (*frente a, jusqu'à, en dehors*)
- no partitive constructions or spatial, temporal or mass nouns (*kilo de, lunes a Viernes, visita a Roma, journées par semaine*)
- no binominal pairs (*dia a dia, instant après instant*)
- no antonyms (*chien sans/avec laisse, personnes avec/sans emploi*)
- no preposition phrase (N1 *à base de, par hasard de*)
- no verb phrase (*mettre N1 en danger, donner N1 à N2*)
- no hybrid forms (*daquela, naquela, deste,…)*

Language	Types	Tokens
French	1063	6991
Spanish	547	10219
Portuguese	6795	58932

Type-Token-Frequencies Datenset3

Quantitative Analysis – first results

Statistical analysis of the factors frequency, productivity and population size*



*My special thanks go to Prof. Dr. Harald Baayen for the statistical analysis of the data

Quantitative Analysis – first results

- Portuguese shows the highest frequency and productivity in the formation of N Prep N constructions and its internal prepositional variation
- in the overall analysis (dataset1), Spanish shows a medium frequency and productivity → the analysis of dataset3 does not support this result
- French appears to be less frequent and productive in dataset1, but this result is contradicted by the analysis of dataset3
- differences in frequency, productivity and population size also appear in the analysis of the prepositional subsets, especially *avec-con-com, à-a-a* and *pour-para-para*
- the results of the prepositional subsets indicate that constructions with the prepositions *avec* and *pour* are less frequent in French (compared to their counterparts *con-com* and *para-para*), while constructions containing *à* are more frequent and productive than its iberoromanic counterparts (*a*)

Qualitative Corpus analysis

Dataset:

manually sorted Datenset3

Language	Types	Tokens
French	1063	6991
Spanish	547	10219
Portuguese	6795	58932

Procedure:

- mapping the prepositional variation on the semantic relations of N1 and N2
- contrastive analysis of specific examples and semantic relations
- comparison of the results with overall frequencies of the constructions and first speaker evaluations

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Qualitative Analyse – Variation *de* and *à/a*

[N1 à/a N2]$_N$ – [N1 de N2]$_N$
a kind of N1 that is defined by N2

[N1 de/à-a N2$_{PURPOSE}$]$_N$
flûtes de/à champagne; fil de/à pêche

forno a/de microondas

[N1 de/à-a N2$_{TYPE/SPECIFICATION}$]$_N$
course à/d'obstacles
freno de/a disco
lampião a/de gás

[N1 de/à-a N2$_{INGREDIENT}$]$_N$
crème au/de citron

biodiesel de/ao óleo

[N1 de/à-a N2$_{BODY\ PART}$]$_N$
fracture au/ du bras
infarto de/del/al miocardio

Only in French:

[N1 de/à-a N2$_{CONTAINER}$]$_N$ → *conteneur de/à déchets, corbeille de/à fruit*
[N1 de/à-a N2$_{MEANS\ OF\ TRANSPORT}$]$_N$ → *course de/à vélo*
[N1 de/à-a N2$_{SEKTOR/BILDUNGSGRAD}$]$_N$ → *étudiants de/au doctorat*

Only in Portuguese:

[N1 de/à-a N2$_{REASON}$]$_N$ → *combate de/a crimes/doenças/incêndio/ obesidade*

see Masini 2009: 263

Qualitative Analyse – Variation *de* and *à/a*

first tendencies

- this variation merely appears in Spanish, only as $N2_{TYPE/SPECIFICATION}$ (*vehículo de/a motor; motor(es) de/a gasolina)*
- as indicated in the quantitative analysis, this variation is more productive in French than in Spanish or Portuguese
- this result is undermined by the comparison with the relative frequency in the overall corpus data as well as first speaker evaluations

- speaker evaluations (as well as literature) highlight a meaning difference:

  flûtes à champagne → *'un verre à usage spécifique'* (Purpose)

  flûtes de champagne → *'un verre qui contient du champagne'* (Content)

→ corpus data include both variants in both meanings! (to be verified by means of experimental investigations)

Qualitative Analysis – Variation *de* and *pour/para*

[N1 *de/pour-para* N2$_{USER}$]$_N$
club d'/pour enfants, collier de/pour chien
club/cuento/juego/ropa de/para niños
abrigo de/para animais; brinquedos de/para crianças

[N1 *de/pour-para* N2$_{PURPOSE}$]$_N$
décoration de/pour mariage/table
agua/alimentos de/para riego/consumo
acessórios de/para cozinha/decoração

[N1 *de/pour-para* N2$_{USER(OBJECT)}$]$_N$
musique de/pour piano
juego de/para pc, música de/para cine
lápis de/para olhos; concerto de/para piano

[N1 *de/pour-para* N2$_{REASON}$]$_N$
acusado do/pelo crime; cirurgia de/para correção

[N1 *de/pour-para* N2$_{TIME}$]$_N$
aluguel de/para férias; hora de/para almoço

[N1 *de/pour-para* N2$_{DETERMINATION}$]$_N$
animais de/para abate/produção, pista de/para caminhada

[N1 *de/por* N2$_{REASON}$]$_N$
prestación(es)/ subsidio de/por desempleo
bônus de/por desempenho

[N1 *de/por* N2$_{INSTRUMENT}$]$_N$
herida(s) de/por arma, internet de/por banda
energia de/por fontes, imagens de/por satélite

Preliminary Conclusion

- Spanish presents the fewest prepositional variation with the fewest types of semantic relations
- Spanish has the strongest tendency to formations with *de*
- French offers more variation with different types of semantic relations than Spanish
- In French, speakers tend the most to rate the use of prepositional variation as ‚incorrect/not acceptable‘
- In some cases, corpus data and speaker evaluations contradict each other in French
- Portuguese shows the most frequent and productive variation that most speakers accept without difficulties

Preliminary Conclusion

- the N Prep N constructions under investigation appear indeed to be formed according to productive word formation rules (or templates) that may underlie variation depending on their semantic properties

- the specific rules of word formation, the preference for a specific prepositional element as well as their productivity vary among the three languages under investigation

- the prepositional element is not a ‚pure linking element, lacking all semantic content', but rather an important element in word formation, whose semantic properties play an important role in the variation of N Prep N constructions as well as the creation of new constructions

- Construction grammar (esp. construction morphology) seems to be an appropriate tool for visualizing first patterns, which will further be investigated by means of psycholinguistic experimentation

References

- Bartning, Inge (1993): La préposition 'de' et les interprétations possibles des syntagmes nominaux complexes. Essai d'approche cognitive. In: Anne-Marie Berthonneau und Pierre Cadiot (Hg.): Les prépositions: méthodes d'analyse. Lille: Presse Universitaires de Lille (Lexique, 11), S. 163–191.
- Benveniste, Émile (Hg.) (1966): Problèmes de linguistique générale. 2 Bände. Paris: Gallimard (Bibliothèque des sciences humaines, 1).
- Berthonneau, Anne-Marie; Cadiot, Pierre (Hg.) (1993): Les prépositions: méthodes d'analyse. Lille: Presse Universitaires de Lille (Lexique, 11).
- Bisetto, Antonietta; Scalise, Sergio (2005): The classification of compounds. In: *Lingue e Linguaggio* 4 (2), S. 319–332.
- Bosredon, Bernard; Tamba, Irène (1991): Verre à pied, moule à gaufres: préposition et noms composés de sous-classe. In: *Langue française* (91), S. 40–55.
- Bouvier, Yves Ferdidand (2000): Definir les composes par opposition aux syntagmes. In: Eric Haeberli und Christopher Laenzlinger (Hg.): Generative Grammar in Geneva, Bd. 1, S. 165–187.
- Buenafuentes de la Malta, Cristina (2006/ 04): Entre la morfología, la sintaxis y el léxico: la delimitaciòn de la composición sintagmática en espanol. Barcelona (VII Congrés de Lingüística General).
- Cadiot, Pierre (1993): À centre deux noms: vers la composition nominale. In: Anne-Marie Berthonneau und Pierre Cadiot (Hg.): Les prépositions: méthodes d'analyse. Lille: Presse Universitaires de Lille (Lexique, 11), S. 193–240.

References (2)

- Di Sciullo, Anne-Marie; Williams, Edwin (1987): On the definition of word. Cambridge, MA (Linguistic inquiry. Monographs, 14).
- Faria, André (2010): Formação de compostos nominais de base livre do PB. In: Maria Lúcia Almeida, Rosangela Ferreira und Diogo Pinheiro (Hg.): Linguística cognitiva em foco: morfologia e semântica do português. Rio de Janeiro: Soluções Editoriais.
- Fradin, Bernhard (2009): IE, Romance: French. In: Rochelle Lieber und Pavol Štekauer (Hg.): The Oxford Handbook of compounding: Oxford University Press, S. 417–435.
- Guevara, Emiliano R. (2012): Spanish compounds. In: Probus. International Journal of Latin and Romance Linguistics 24 (1), S. 175–195.
- Kornfeld, Laura Malena (2009): IE, Romance: Spanish. In: Rochelle Lieber und Pavol Štekauer (Hg.): The Oxford Handbook of compounding: Oxford University Press, S. 436–453.
- Lieber, Rochelle; Scalise, Sergio (2007): The lexical integrity hypothesis in a new theoretical universe. In: Geert Booij, Luca Ducceschi, Bernhard Fradin, Emiliano R. Guevara, Angela Ralli und Sergio Scalise (Hg.): On-line Proceedings of the Fifth Mediterranean Morphology Meeting. Fréjus, 15-18.09.2005, S. 1–25.
- Masini, Francesca (2009): Phrasal lexemes, compounds and phrases. A construcionist perspective. In: Word Structure 2 (2), S. 254–271.
- Masini, Francesca; Scalise, Sergio (2012): Italian compounds. In: Probus. International Journal of Latin and Romance Linguistics 24 (1), S. 61–91.

References (3)

- Masini, Francesca; Thornton, Anna (2007): Italian VEV lexical constructions. In: Geert Booij, Angela Ralli und Sergio Scalise (Hg.): Morphology and Dialectology. Sixth Mediterranean Morphology Meetings. Idaca (6), S. 148–189.
- Moyna, María Irene (2011): Compound words in spanish. Theory and history. Amsterdam/ Philadelphia: John Benjamins (Current Issues in Linguistic Theory. Amsterdam Studies in the Theory and History of Linguistic Science, 316).
- Rackow, Ulrike (1994): Vent d'ouest - vent du nord/ hombre de campo - hombre del campo. Zum Problem des 'Binnenartikels' im Französischen und Spanischen. Tübingen: Gunter Narr Verlag (402, Tübinger Beiträge zur Linguistik).
- Rainer, Franz (2016): Italian. In: O. Peter Müller, Ingeborg Ohnheiser, Susan Olsen und Franz Rainer (Hg.): Word Formation. An International Handbook of the Languages of Europe, Bd. 4. Berlin, Boston: Mouton de Gruyter, S. 2712–2731.
- Rainer, Franz (2016): Spanish. In: O. Peter Müller, Ingeborg Ohnheiser, Susan Olsen und Franz Rainer (Hg.): Word Formation. An International Handbook of the Languages of Europe, Bd. 4. Berlin, Boston: Mouton de Gruyter, S. 2620–2640.
- Rio-Torto, Graça; Ribeiro, Sílvia (2009): Compounds in portuguese. In: Lingue e Linguaggio 8 (2), S. 271–291.
- Scalise, Sergio; Magni, Elisabetta; Bisetto, Antonietta (Hg.) (2009): Universals of language today. Rotterdam: Springer Science + Business Media B.V. (Studies in Natural Language and Linguistic Theory, 76).
- Villoing, Florence (2012): French compounds. In: Probus. International Journal of Latin and Romance Linguistics 24 (1), S. 29–60