

GL2013

Proceedings of the
**6th International Conference on
Generative Approaches to the Lexicon**
Generative Lexicon and Distributional Semantics

Edited by
Roser Saurí, Nicoletta Calzolari, Chu-Ren Huang, Alessandro Lenci,
Monica Monachini, James Pustejovsky

September 24-25, 2013
Pisa, Italy

ISBN 978-1-937284-98-5



Organizers:

Nicoletta Calzolari Istituto di Linguistica Computazionale “Antonio Zampolli”, Pisa, Italy
James Pustejovsky Brandeis University, Waltham, MA, USA

Chairs:

Chu-Ren Huang The Hong Kong Polytechnic University, Hong Kong
Alessandro Lenci Università di Pisa, Pisa, Italy
Monica Monachini Istituto di Linguistica Computazionale “Antonio Zampolli”, Pisa, Italy
Roser Saurí Barcelona Media, Catalonia, Spain

Local Organizer:

Sara Goggi Istituto di Linguistica Computazionale “Antonio Zampolli”, Pisa, Italy

Sponsors:



Endorsed by:



Program Committee

Marco Baroni	University of Trento
Olga Batiukova	Universidad Autonoma de Madrid
Núria Bel	Universitat Pompeu Fabra
Sabine Bergler	Concordia University
Gemma Boleda	University of Texas
Pierrette Bouillon	ETI/TIM/ISSCO, University of Geneva
Nicoletta Calzolari	ILC-CNR
Philipp Cimiano	University of Bielefeld
Ann Copestake	University of Cambridge
Laurence Danlos	Universite Paris 7
Stefan Evert	University of Erlangen
Christiane Fellbaum	Princeton University
Shu-Kai Hsieh	National Taiwan University
Chu-Ren Huang	The Hong Kong Polytechnic University
Nancy Ide	Vassar College
Hitoshi Isahara	Toyohashi University of Technology
Elisabetta Jezek	Università di Pavia
Kyoko Kanzaki	Toyohashi University of Technology
Adam Kilgarriff	Lexicography MasterClass Ltd
Alessandro Lenci	University of Pisa - Department of Linguistics
Bernardo Magnini	FBK
Louise McNally	Universitat Pompeu Fabra
Monica Monachini	ILC-CNR
Seungho Nam	Seoul National University
Fiammetta Namer	ATILF-CNRS, University of Nancy
Sebastian Padó	University of Heidelberg
Martha Palmer	University of Colorado
Massimo Poesio	University of Trento
James Pustejovsky	Brandeis University
Valeria Quochi	Instituto Di Linguistica Computazionale "Antonio Zampolli"
German Rigau	UPV/EHU
Anna Rumshisky	University of Massachusetts
Magnus Sahlgren	Gavagai AB
Roser Saurí	Barcelona Media
Zuoyan Song	Beijing Normal University
Laure Vieu	Institut de Recherche en Informatique de Toulouse
Piek Vossen	Vrije Universiteit
Alessandra Zarcone	University of Stuttgart


 The logo for Diacemto, featuring a stylized 'D' followed by the word 'iacemto' in a lowercase, sans-serif font.


Preface

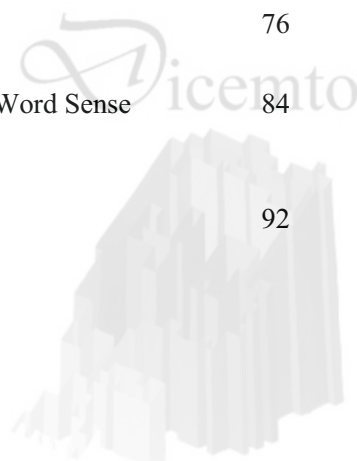
The papers in this volume represent some of the most recent and exciting work being carried out both within the framework of Generative Lexicon and related approaches to the lexicon and lexical resources. With the recent emphasis in natural language processing on the development of machine learning algorithms, it has become even more important for computational linguists to work on the development of linguistically informed lexical resources, for use in the annotation of corpora and creation of gold standard data for training, as well as the collation of larger theoretical datasets for investigating linguistic phenomena in greater detail and sophistication. These works contribute to this trend as well as to the further development of the mechanisms within GL for describing and explaining semantic and lexical phenomena in language.

The GL2013 Organizers and Chairs



Table of Contents

Program Committee	ii
Preface	iii
Talks	
<i>Invited Talks</i>	
<hr/>	
Dynamic Event Structure and Habitat Theory <i>James Pustejovsky</i>	1
Metaphor and Qualia: Embodiment or Eventuality? <i>Chu-Ren Huang, Kathleen Ahrens and Francesca Quattri</i>	11
<i>Day 1</i>	
<hr/>	
To Coerce or Not to Coerce: A Corpus-based Exploration of Some Complement Coercion Verbs in Chinese <i>Chan-Chia Hsu and Shu-Kai Hsieh</i>	13
Towards the automatic classification of complex-type nominals <i>Lauren Romeo, Sara Mendes and Núria Bel</i>	21
Qualia Relations in Metaphorical Noun-Noun Compounds <i>Zuoyan Song and Qingqing Zhao</i>	29
From Glosses to Qualia: Qualia Extraction from Senso Comune <i>Tommaso Caselli and Irene Russo</i>	37
Expanding VerbNet with Sketch Engine <i>Claire Bonial, Orin Hargraves and Martha Palmer</i>	44
Primitives of Events and the Semantic Representation <i>Hongzhi Xu and Chu-Ren Huang</i>	54
Generative Lexicon Theory and Linguistic Linked Open Data <i>Fahad Khan, Francesca Frontini, Riccardo Del Gratta, Monica Monachini and Valeria Quochi</i>	62
<i>Day 2</i>	
<hr/>	
Disambiguation of Basic Action Types through Nouns Telic Qualia <i>Irene Russo, Francesca Frontini, Irene De Felice, Fahad Khan and Monica Monachini</i>	70
Class-based Word Sense Induction for dot-type nominals <i>Lauren Romeo, Héctor Martínez Alonso and Núria Bel</i>	76
Mixing in Some Knowledge: Enriched Context Patterns for Bayesian Word Sense Induction <i>Rachel Chasin and Anna Rumshisky</i>	84
Informativeness Constraints and Compositionality <i>Olga Batiukova and James Pustejovsky</i>	92



Extended Generative Lexicon <i>Sumiyo Nishiguchi</i>	101
Features of Verb Complements in Co-composition: A case study of Chinese baking verb using Weibo corpus <i>Yu-Yun Chang and Shu-Kai Hsieh</i>	106
A Lexico-Semantic Analysis of Chinese Locality Phrases - A Topic Clustering Approach <i>August F.Y. Chao and Siaw-Fong Chung</i>	115
Author Index	125
Keyword Index	126



Informativeness Constraints and Compositionality

Olga Batiukova

Spanish Philology Department
Autonomous University of Madrid
Cantoblanco, Madrid, Spain
volha.batsiukova@domain

James Pustejovsky

Computer Science Department
Brandeis University
Waltham, Massachusetts, USA
jamesp@cs.brandeis.edu

Abstract

In this paper we examine the role that compositional mechanisms and lexical semantics play in the determination of *informativeness* at the phrasal and clausal level. While the computation of the “relevance” of an utterance is largely determined by pragmatic factors (such as quantity), we argue that phrasal informativeness can, in many cases, be computed compositionally and independently of pragmatics. To illustrate this, we focus on the well-documented contrast between predicative and derived participial modificational constructions in English (*build a house* results in well-formed sentences, while **a built house* does not). In our analysis, informativeness within an NP is computed in terms of minimal model generation (Blackburn and Bos, 2008), using the semantics associated with the qualia of the head noun; that is, modification is informative whenever a qualia value is not satisfied in all models.

1 Introduction

This paper studies the contrast in acceptability of certain past participle-noun (PP-N) modification constructions and their corresponding verb-noun predicates (V-N), as illustrated in (1):

- (1) a. buy a ticket vs. *a bought ticket
- b. eat a sandwich vs. *an eaten sandwich
- c. feel sympathy vs. *a felt sympathy
- d. give an answer vs. *a given answer
- e. hear a noise vs. *a heard noise
- f. make a mistake vs. *a made mistake
- g. play the piano vs. *a played piano
- h. read the newspaper vs. *a read newspaper
- i. win the prize vs. *a won prize
- j. write a book vs. *a written book
- k. see the movie vs. *a seen movie

This is surprising, given the semantic similarity between verb argument selection and the corresponding modification operation. For example, both elements in the pairs below are well-formed.

- (2) a. paint a house vs. a painted house

- b. spill the milk vs. the spilled milk
- c. poison the food vs. poisoned food

The question immediately arises as to why such a distinction in grammaticality should exist, as well as what the constraints affecting the well-formedness of these constructions might be. This topic has been approached from two different perspectives in the literature, which we review briefly before presenting our proposal: (a) an aspect-based approach; and (b) a pragmatically-determined informativeness approach.

According to the aspect-based approach (argued for in Bresnan (1995), Langacker (1991), Valin (1990), Embick (2004)), a PP-N construction is grammatical if the participle denotes the resultant state of the verb from which it is derived. However, most unacceptable combinations in (1) meet this requirement: they are either achievements (*given answer, made mistake*) or accomplishments (*bought ticket, eaten sandwich, written book*).

Grimshaw and Vikner (1993) introduce an additional requirement in their study of obligatory adjuncts in passives: each of the subevents of the event structure of the verb has to be identified by an argument. The only overt argument of PP-N constructions is usually the theme, which is involved in both subevents if the event is complex (e.g., *the ruined shirt* is an accomplishment composed of a process and a resultant state, both of which are related to *shirt*). Creation predicates are one exception, because the theme is related to the resultant state only (the object does not exist until the event is completed). This is why a second element, an adjunct, is needed to identify the process subevent, e.g., *an expertly written book*.

Under this assumption, in the rest of the examples in (1) one argument should be enough to guarantee the acceptability of the construction, which is obviously not what we get.¹ Grimshaw

¹See Jung (1997) and Ackerman and Goldberg (1996) for a detailed criticism of aspect-based approaches.

and Vikner (1993) do mention an alternative approach to this issue in the conclusion of their study, where they suggest that the obligatory adjunct phenomenon is a matter of satisfying the requirement that one ‘say something’.² This ‘say something’ requirement has been interpreted in Jung (1997) as a general *pragmatic* condition on presupposition and assertion in passives: “The predicate must assert more than what is presupposed by the subject”. As definite NP subjects bear an existential presupposition, the reference to their creation violates the *Say Something Condition*. Any adjunct providing new information will qualify as compulsory in this situation. Compare the examples in (3):

- (3) a. *A house* was built / * *This house* was built / *This house* was built *to our specification*.
 b. *A picture* was taken / * *This picture* was taken / *This picture* was taken *to my liking*.

Following similar assumptions, the account by Ackerman and Goldberg (1996) is also pragmatically motivated. It is based on the Gricean maxim of Quantity (‘make your contribution as informative as required for the current purposes of exchange; do not make your contribution more informative than is required’) and Horn’s R-principle (‘make your contribution necessary; say no more than you must’) (Levinson (2000) and Horn (1996)). They claim that “adjectival past participles (APP) can only occur if they are construable as predicating an informative state of the head noun referent”. This claim is based on two constraints:

1. **Non-redundancy constraint:** If the referent of the head noun, N, implies a property P as part of its frame-semantic or encyclopedic knowledge, then the APP is not allowed to simply designate P; it must be further qualified.
2. **Paradigmatic Informativeness constraint:** An APP phrase is not felicitous if it is based on a superordinate level verb which contrasts with semantically more specific predicates (troponyms).

The non-redundancy constraint clearly accounts for cases in (1): all the newspapers are meant to be read, sympathy only arises when it is felt somehow, and so on. The addition of an adverb (4), an adjectival or nominal modifier ((5) and (6)), as well as certain morphological elements (derivational affixes, as in (7)) makes the property denoted by the participle more specific and renders the whole construction informative:

- (4) a. *bought ticket vs. {recently / illegally / their already / the most} bought ticket
 b. *eaten sandwich vs. {quickly / half / partially} eaten sandwich

²D. Pesetsky, p.c.

- c. *felt sympathy vs. {suddenly / heart / deep / instantly} felt sympathy
 d. *given answer vs. {previously / frequently / commonly / the above} given answer
 e. *heard noise vs. {barely / abnormally / repeatedly} heard noise
 f. *made mistake vs. {stupidly / easily / often / widely} made mistake
 g. *played piano vs. {beautifully / passionately / badly / gently} played piano
 h. *read newspaper vs. {carefully / widely / the most} read newspaper
 i. *won prize vs. {easily / rightly / fraudulently} won prize
 j. *written book vs. {well / poorly / engagingly / intelligently / newly / vividly} written book
 k. *seen movie vs. {last / little / never before / rarely} seen movie
 (5) a. *manufactured aircraft vs. {contemporary / American} manufactured aircraft
 b. *published books vs. {recent / foreign} published books
 (6) a. ??trained people vs. science-trained people
 b. *shaped fish vs. angle-shaped fish
 (7) ??arranged rendezvous vs. pre-arranged rendezvous

The Paradigmatic Informativeness constraint is designed to explain the cases in (8), where a verb denoting a particular manner of performing the action is preferred to the less specific superordinate verb:

- (8) a. *cut meat vs. sliced / chopped meat
 b. *told secret vs. disclosed / confessed secret
 c. *given funds vs. donated / sacrificed funds

Note, however, that some of these examples are odd even if we add adverbial modifiers:

- (9) a. ?quickly told secret
 b. ?recently given funds
 c. ?secretly taken shirt

While both constraints proposed in Ackerman and Goldberg (1996) seem to be on the right track, the notions they are based on (frame-semantic and encyclopedic knowledge) are left rather vague. Many things can be ascribable to encyclopedic knowledge. As for frame-semantic content, this can extend to an unrestricted repertoire of specific semantic and situational parameters (roles and otherwise). This vagueness and unrestrictedness makes it difficult to formalize both constraints and how to apply them .

In a move to remedy this vagueness, Goldberg and Ackerman (2001) propose a more general requirement for modification and predication constructions: they must be informative in the conversational context. One way the utterance can be informative is by containing a focus (provided by negation, modality, tense, aspect, adjunct, indefinite subject, etc.) that conveys something non-presupposed.

- (10) a. The house was built.
 b. The house was not built. NEGATION
 c. The house {should/might} be built. MODALITY

- d. The house {will be / has been} built.
TENSE/ASPECT
e. The house was built {last year}. ADJUNCT
f. A house has been built. INDEFINITE

As the adjectives and participles in modification constructions have less linguistic information associated with them than verbs (there is neither tense nor modality, and the array of aspectual interpretations is very limited), it is more difficult to provide a focus for a successful assertion (relative acceptability is indicated by ' > '):

- (11) a. #This house was built. > #a built house
b. #That book was read. > #the read book
c. #The television program was watched. > #the watched program

While we acknowledge that much of the “informativeness” of lexical choice in an utterance can be determined only after most contextual variables are already fixed, we argue that there are *compositional aspects* to the calculation of informativeness that have not been adequately appreciated.

In the remainder of the paper, we show that a significant part of what is called “informativeness” can be accounted for compositionally. Following Konrad (2004) and Blackburn and Bos (2008), we utilize minimal model generation as part of the compositional computation, where we assume that a linguistic expression should be *consistent* within a discourse and *informative* relative to what is known. In model-theoretic terms, *consistent* means ‘satisfied in at least some models or situations’ (cf. the formal definition in the next section). Within the compositional construction of an utterance itself, we can compute consistency as type satisfaction (Pustejovsky, 2013), as assumed within typed functional languages. An expression is informative on the other hand, if it is ‘not satisfied in all models and with all assignments’. Our treatment of informativeness is based on the semantics provided by the *qualia*, a structured representation of the meaning parameters encoded by lexical items (Pustejovsky, 1995): that is, whenever a *qualia* value is not attested in all possible situations involving a given expression (i.e. not satisfied in all models), the expression will be judged informative. We outline the basic ideas behind this approach in the section below.

2 General Hypothesis and Predictions

Our starting assumption relates to the definition of semantic predication and argument selection. We believe that the contrast in acceptability between predication and modification constructions involving the same elements (cf. the examples in (1)) can be better accounted for if we assume that

both constructions are instances of semantic predication. The main difference is that in a V-N construction the verb is the predicate projecting the argument structure, imposing selectional requirements on its arguments, while in a modification construction the noun is the head, yet it projects its argument structure as well. A brief motivation of this step is in order.

Verbs and deverbal nominals are traditionally considered as prototypical relational items bearing the predicative force: they select for certain kinds of elements (arguments) compatible with them, which complete and specify their meaning. Chomsky (1993), Goldberg (1998), Dowty (1979), Croft (2005), among others, assume a verb-centered bias toward how arguments are identified in the phrase and sentence, be they verbs or relational nouns.

As is well known, the Generative Lexicon focuses to open up the channel of relation identification and argument selection through the introduction of non-verb based argument associations, i.e., the *Qualia Structure roles* associated with the nouns constituting arguments and adjuncts in the sentence. The four parameters encoded in the Qualia Structure are AGENTIVE (factors involved in the origin or creation of entities and events, such as *build* for *house*), CONSTITUTIVE (internal constituency of the whole, such as constituent parts of material entities), FORMAL (the distinctive features of entities, such as spatial orientation, size, shape, dimensionality, color, etc., and the taxonomic relations, e.g., a *house* is a *building*), and TELIC (purpose and function of entities and events, such as *reading* for *book*).

The Qualia Structure can be regarded as similar in many respects to the Argument Structure for verbs. In a fashion similar to Argument Structure realization, the Qualia roles do not need to be expressed overtly in order to be accessible for interpretation. Just as the verb *eat* presupposes that its direct object denote a kind of food even when not overtly expressed, nouns may encode “hidden” relations along with unexpressed arguments; e.g., the relation of inalienable possession denoted by the noun *hand*, as being a part of a *body*, to mention just one of the syntactically relevant semantic relationships. Artifactual nominals, in addition, refer to the event which brought them about and to the activity they are meant for: e.g., *house* presupposes a creation event, as well as a functional value associated with its purpose.

As we anticipated at the end of the previous section, qualia are crucially involved in the compositional calculation of consistency and informativeness of linguistic expressions. A consistent utterance describes a realizable situation, that is, representable as a first-order formula satisfied in at least some models. All the arguments must be consistent with the predicate, in the sense of “semantically compatible” (e.g., *male* is consistent with the semantics of the noun *bachelor*, while *married* is not). This applies to both arguments (in the strict sense of the term) and adjuncts. Inconsistent combinations should not be present in natural data. Informative utterances are a subset of consistent utterances, whose denotation is ruled out in at least some situations. Hence, while both *male* and *funny* are consistent with *bachelor*, only *funny bachelor* is an informative phrase, since not all bachelors are funny.

In typed functional languages, consistency is defined as *type satisfaction*: the argument must have the type required by the predicate or function. In GL, four predicative compositional mechanisms have been identified: *type matching* or *pure selection*, *accommodation*, *coercion by introduction* and *coercion by exploitation* (Pustejovsky, 2011; Asher and Pustejovsky, 2013). *Type matching* takes place when the type required by the verb is directly satisfied by the argument (e.g. *read a book*: *book* is $phys \bullet info$ and *read* is $phys \bullet info \rightarrow (e \rightarrow t)$).³ *Accommodation* allows combining a predicate with an argument whose hypernym satisfies its selectional requirements through type inheritance (e.g. *the beer spoiled*: *spoil* is $phys \otimes_T \tau \rightarrow t$, and it can be combined with *beer*: $liquid \otimes_T drink$, because $liquid \subseteq phys$ and $drink \subseteq \tau$). *Coercion* mechanisms are activated when the type a function requires is imposed on the argument type. In these cases, the qualia act as type shifting operators, allowing an expression to satisfy new typing environments through *introduction* or *exploitation*. In *enjoy a coffee*, for example, both mechanisms are consecutively activated: *enjoy* needs a direct object typed as *event*, and *coffee* must first be wrapped with the type *event* through introduction (*coffee:event*), and af-

³The following notation is used in this paragraph: τ and T refer to the telic role, and \bullet (the dot) and \otimes (the tensor) are type constructors. The dot builds the *dot objects*, such as *book* above, and the tensor introduces agentive and telic information to the head type to derive artifactual types, e.g. *beer*.

terwards the value in the telic role of *coffee* is exploited to turn it into *coffee:drink* event.

To make our computation of consistency and informativeness more explicit, we adopt a strategy of *model generation* (Blackburn and Bos, 2008; Konrad, 2004).⁴ The consistency of an expression, $\lambda x[F(x)](A)$, after function application of F over A , can be checked by determining whether the set of first-order formulas resulting from the application are satisfiable (i.e., there is a model \mathcal{M} corresponding to this set). The informativeness of a function application can be similarly defined: a function application, $\lambda x[F(x)](A)$, is informative if and only if the set of first-order formulas resulting from the application is not satisfied in all models, \mathcal{M}_i .

We are now in position to take a closer look at the informative contribution of consistent arguments to the semantics of the resulting expression. Clearly, *non-required* arguments (adjuncts) are always informative, since they contribute additional information not deducible from the predicate meaning. *Required* arguments are a necessary part of the logical form of the predicate, but they may be left unexpressed in syntax for different reasons, due to anaphoric binding for example. Here we are interested in required arguments whose semantic content is incorporated in the predicate, i.e. the *default* arguments of the classical GL (Pustejovsky, 1995). These arguments can only appear when their denotation is informative with respect to the head, i.e., when there is a model and assignment where the resulting expression is not true. When uninformative, they are left unexpressed or *shadowed* by the predicate.

Shadowed arguments are assigned a very general interpretation, which has the same level of specificity of the semantic type imposed by the predicate. For instance, the default argument of *eat* is interpreted out of context as ‘something eatable’ (indefinite and non-specific) rather than a specific kind of food, and the default way of coming into being of *a sheep* is *to be born* rather than *cloned*.

The asymmetry in informativeness-determined acceptability of V-N predicative constructions and PP-N modification constructions emerges when the nominal argument is required by the verb and is informative with respect to it, but the verb (its

⁴We discuss the details of the mechanism elsewhere, Pustejovsky and Batiukova (forthcoming).

participial form) is a default argument of the noun, and it fails to be informative: *eat a sandwich* is informative because many other things can be eaten (i.e., *sandwich* is more specific than the type selected by *eat*, which is FOOD). At the same time, *eaten sandwich* is uninformative because all the sandwiches are meant to be eaten: *eat* is the default argument (or *default telic*, in terms of qualia) of *sandwich*, it is uninformative with respect to the nominal head and therefore must be shadowed.

Even though the semantic mechanisms underlying predication and modification are different, we suggest that the same compositional principles are at play as far as consistency and informativeness of the argument with respect to the syntactic head is concerned. Predication is typically viewed as function application, whereby the predicate is applied to an argument in order to obtain a truth value. In the classical GL, modifying adjectives have been analyzed as typed functions applied to a particular quale of the head noun by means of *selective binding* or *subselection*. For example, *good* targets specifically the event description encoded in the telic role, and *long* can refer to one of the dimensions of a physical object or to the duration of the event referred to in one of the qualia of the head noun:

- (12) a. *good teacher*: a teacher who teaches well; *a good knife*: a knife that cuts well
 b. *long shadow*: a shadow having greater extension than usual; *long vowel*: a vowel whose pronunciation has a certain duration

Modifications introduced in recent versions of the theory suggest that the selectional mechanisms involved in verbal constructions can be applied to adjectival modification as well. In both kinds of constructions, type adjustment is guided by the *Head Typing Principle*, according to which the typing of the head must be preserved in any composition rule (Asher and Pustejovsky, 2013).

In both modification and predication constructions, the argument must be informative with respect to the syntactic head, hence the degree of informativeness of the construction is crucially determined by the mechanism involved in the combination of both elements: *type matching* gives rise to expressions with a very low degree of informativeness (which can even be zero or nonexistent), since the semantics of the argument is largely included in the meaning of the predicate. The compositional mechanisms of *accommodation* and *introduction* are always informative, the former less than the latter, since the argument is basically a

subtyped version of the required type. As far as coercion is concerned, *introduction* is always informative, since the argument is wrapped with a new type, not entailed by argument's semantics. Note that *exploitation* is never inherently informative, since the semantic content is entailed by the argument's semantics.

From what has been said in this section, we can make the following generalizations and predictions, which will be tested in the following sections:

- The degree of informativeness of the PP-N combinations must be determined compositionally: the same modifier can be redundant or informative depending on the semantics of the head noun.
- Acceptable PP-N combinations cannot refer to the default qualia values of the head noun, such as physical parameters or internal constituency of the denoted entity. In addition, artifact-denoting nouns should not be compatible with modifiers referring to default function or origin.
- Whenever a priori uninformative PP-N combinations appear in natural data, this is due to the intervention of one of the rescue mechanisms: (1) the *default informative mechanism* is the contrastive reading, which presupposes a binary partition of the set of discourse elements (e.g. a BUILT house as opposed to non-existent or partially built houses)⁵ and (2) the presence of an additional modifier attached to the construction, as in (4).

3 Source of data

The data analyzed in this study were extracted from the enTenTen12 corpus (using Word Sketch, cf. Kilgarriff et al. (2004)) and supplemented by introspective data. The search queries were defined for past participles followed by a noun. Two types of sequences were filtered out in the initial and the final position, respectively: the auxiliary *have*, to discard the present perfect forms, and postponed nouns, which give rise to compounds (as in *associated e-mail address*). Two types of forms were obtained this way: adjectival and participial deverbal *-ed* forms (e.g., *baked*, *broken*, *employed*, *seen*, etc.), and denominal adjectival forms ending in *-ed*, which will be referred to as *pseudo-participles*: *winged* (as in *winged aircraft*), *sanded* (as in *sanded dust*), etc. The decision of including deverbal adjectives along with true participles was motivated by the fact that the

⁵A reviewer points out that the possibility of contrastive interpretation for uninformative constructions indicates that pragmatics ultimately determines whether an expression is informative or not. We believe that this is not the case, since lexical semantics and pragmatics operate on different levels: pragmatics can not explain why certain word combinations (e.g. *eaten sandwich*) are uninformative, because it has no access to the internal structure of words, but it can make them acceptable in context by expanding the universe of discourse (e.g. by including the non-consumed sandwiches therein).

limit between these two categories is not clearly defined in many cases. As a matter of fact, the same item in a similar distribution was classified in enTenTen12 as a past participle in some instances and as an adjective in others (cf. *illustrated, damaged, introduced*, etc.). We also included the denominal derivatives, since the exact categorial nature of the prenominal modifier is not crucial for us. The main goal is to identify the constraints on informativeness operating in modification constructions.

In this study we compare nouns differing with respect to two sets of features, *natural / artificial* and *count / mass*: *water, dust, sand* (natural, mass), *wine* (artifact, mass), *tree* (natural, count), *aircraft* (artifact, count). A total of 3350 PP-N pairs were extracted for *tree*, 777 for *sand*, 1241 for *dust*, 9350 for *water*, 3098 for *aircraft* and 7743 for *wine*. The annotation of the extracted pairs involved judging the grammaticality of the PP-N constructions without additional modifiers (of the kind illustrated in (4)-(7)), annotating the PP modifiers as default and non-default, and identifying the qualia roles they bind. For space reasons, only a small sample of all the attested PP-N combinations is explicitly referred to in what follows. We are particularly interested in the behavior of the PPs that bind one of the qualia roles, in order to test the hypothesis of *qualia informativeness* as formulated above: the modifier can not refer to the default qualia values of the head unless subtyped or given a contrastive reading.

4 Qualia Informativeness: Formal and Constitutive

All the nouns in our sample are compatible with PPs referring to the distinguishing physical properties of the denoted entities, whenever these properties are not default. *Colored* and *shaped* refer to a default attribute of most physical objects, therefore they need to be subtyped to be informative:

- (13) a. *(deeply / garnet / beautifully) colored wine
 b. *(naturally / white, brightly) colored sand
 c. *(red / mud / orange / non-) colored dust
 d. *(green / brightly / unusually) colored tree
 e. *(white / vibrantly / oddly) colored aircraft
 f. *(nicely / strangely / beautifully) shaped tree

If there is no modifier, *colored* is interpreted as ‘artificially or unusually colored’ for natural entities (*sand, dust, and tree*). This is the only possible interpretation of *colored water*, too, but for a different reason: *water* lacks the color attribute, therefore it is always informative.

- (14) a. For this you may need colored sand
 b. small quantities of what looks like colored dust

- c. consider buying a colored tree and decorating it with dazzling lights
 d. Allow each egg to stay in the colored water for increasingly more time

The same can be said about PPs referring to the internal constituency of both naturals and artifactuals: default constitutive attributes are shadowed unless subtyped:

- (15) a. *(suitably / properly / similarly / specially / ADS-B) equipped aircraft
 b. *(wide / narrow) bodied aircraft
 c. *(full / light / heavy) bodied wine⁶
 d. *(large / goof / coarse) grained sand
 e. *(un- / well / strongly / firmly) rooted tree
 f. *(thickly / fully / sparsely / low) branched tree

The default argument can only appear unmodified if it yields a contrastive interpretation. The following example, for instance, can only be interpreted as ‘branched tree as opposed to trees without branches’:

- (16) in the shape of a branched tree

Combinations with non-default constitutives are informative, hence acceptable: not all aircrafts have wings (e.g. the helicopters do not) and not trees have leafs (e.g. coniferous trees do not).

- (17) a. winged aircraft
 b. leafed tree

5 Qualia Informativeness: Agentive

Markedness for origin and function is a prominent part of the lexical semantics of artifactuals as opposed to natural types: artifacts are entities created with a specific purpose or as a result of a purpose-driven activity. The default agentive value encoded in the lexical entry of artifactual nominals must be further specified in order to yield an informative construction:

- (18) a. *(poorly / locally / well / excellently / sustainably / your own) made wine⁷
 b. *(mass / commercially / exclusively / locally) produced wine
 c. *(Soviet / commercially) made aircraft
 d. *(newly / technically / recently / fully) developed aircraft
 e. *(commercially / domestically) produced aircraft

The same holds for metonymic interpretations, as in (19): strictly speaking, wine does not grow, but the grapes do (i.e., *grown* does not bind the agentive of *wine* directly, but through consecutive applications of *exploitation* of the agentive: wine is made of grapes or grape juice, which in turn come into existence by the process of growing).

- (19) *(locally / organically) grown wine

⁶When applied to *wine*, *bodied* does not refer to its internal structure or ingredients. Rather, it describes the taste.

⁷*Made wine* can refer to a specific kind of alcoholic beverage, different from wine.

When the participle describes a specific, non-default way of creating the artifact, the combination is informative:

(20) *Graham* defines this as a crafted wine.

Unlike artifacts, natural kinds are underspecified for origin. However, it can be referred to explicitly with the same restrictions as for artifacts.

(21) a. air-born dust
b. melted water
c. *(farm / seed / field / container) grown tree ⁸

When naturally-occurring entities are produced artificially, the reference to origin becomes informative (by the mechanism of *introduction*, which always generates informative combinations, as argued in section 2):

(22) a. {manufactured / produced} sand
b. produced water
c. {ready / badly} made tree
d. {created / planted} tree

6 Qualia Informativeness: Telic

Following our hypothesis stated above, the activity associated with the telic quale of an object, when used in the PP-N construction, should be (modally) uninformative relative to the head.

(23) a. *(locally) eaten meat
b. *(rarely) driven car
c. *(seldom) watched film

We can account for this by constructing a minimal modal model, capturing the modal subordination inherent in the Telic value. Minimal model construction can reflect the modal subordination inherent in the telic role, following Blackburn and Bos (2008).⁹ Informally, this says that the bare participial modifiers in (23) are uninformative, relative to the minimal modal models generated from the telic values for each of the respective head nouns. According to this analysis, artifact-denoting nouns in general should not be compatible with default telic arguments. Again, the prediction seems to be borne out, as seen in (24).

(24) a. *(commonly / widely / most often) drunk wine
b. *(remotely / carelessly / frequently / previously) flown aircraft

Natural kinds are underspecified for function (the telic role). However, they can be routinely recategorized to refer to some kind of conventionalized use, as seen in *drinking water*, *edible fruit*, etc. These combinations are possible due to *qualia introduction*, and hence their informativeness. In (25), *used water* and *used sand* are interpreted as ‘used before for human activity, not clean’. *Used tree*, in turn, refers to the Christmas tree when there is no modifier:

⁸This example is acceptable without modifier if *grown* refers to the size of the tree rather than to its origin.

⁹See Pustejovsky and Batiukova (forthcoming) for more details.

(25) a. The used water is fed back into the source for re-heating.
b. There is potential for used sand to contain toxic or harmful ingredients.
c. Make it a resolution this new year to keep your used tree out of a landfill.

Our hypothesis predicts an inverse relationship between the degree of lexical-semantic specificity of different groups of nominals and the range of modifiers they are compatible with: since the artifactual types have more lexical-semantic information associated with them than the naturals, they are expected to reject a greater number of modifiers due to the informativeness constraint. This prediction can be tested statistically by calculating what percentage of PP-N combinations require an additional modifier in order to be informative. Although a much larger data sample is needed to get reliable results, we can say that this prediction is borne out for the six nominals examined here. The percentage of PP-N pairs with an additional modifier is higher when the head is an artifactual type: *tree*-31.43%, *sand*-31.02%, *dust*-22.08%, *water*-19.05%, *aircraft*-44.19%, *wine*-34.94%.

7 Conventionalized Attributes

A significant portion of what we know about events and their associated participants is not encoded linguistically (i.e., it does not affect the syntactic behavior of lexical items) and is not directly encoded in the lexical structures (the argument structure, the event structure or the qualia structure). Some aspects of such information, however, may be prominent both cognitively and statistically. This is what is called *conventionalized attributes* in Pustejovsky and Jezek (2008) or *Generalized Event Knowledge* in a recent trend in psycholinguistics (McRae and Matsuki, 2009). Here are some examples:

(26) a. *(moderately) priced wine
b. *(high / top) rated wine
c. *(full / heavy / light) bodied wine
d. *(strategically / conveniently) placed tree
e. *(well / professionally / badly) maintained aircraft

These attributes seem to behave similarly to true arguments: whenever a conventionalized attribute is entailed by the semantics of the head noun, it must be shadowed unless subtyped.

8 Data Summary

The following tables summarize the cases discussed in sections 4-6, with some additional corpus examples added for illustrative purposes. Even though only a small sample of all the analyzed data is reflected here, the validity of the overall predicted pattern has been confirmed in a thorough manual data analysis: default modifiers can

only appear without an adjunct when the sentence has a contrastive reading or as a consequence of coercion by introduction.

The following types of modifiers are included in the second column for all the qualia roles ('F/C' means 'formal/constitutive', 'A' 'agentive', and 'T' 'telic'): modified defaults, unmodified defaults with a contrastive or coerced interpretation, and non-default subtyped modifiers.

Qualia	PP Modifier	Examples
F/C	Modified default	<i>colored, shaped, rooted, branched, formed, headed, crowned</i>
	Contr./C-E default Subtyped	<i>colored, branched, curved leafed, unrooted</i>
A	Modified default	<i>grown, made, created, planted, cultivated, cloned</i>
	Contr./C-E default Subtyped	
T	Modified default	<i>used, harvested</i>
	Contr./C-E default Subtyped	

Table 1: *Tree*

Qualia	PP Modifier	Examples
F/C	Modified default	<i>colored, grained</i>
	Contr./C-E default	<i>colored</i>
	Subtyped	<i>bleached</i>
A	Modified default	<i>manufactured, produced, excavated, eroded, obtained</i>
	Contr./C-E default Subtyped	
T	Modified default	<i>used</i>
	Contr./C-E default Subtyped	

Table 2: *Sand*

Qualia	PP Modifier	Examples
F/C	Modified default	<i>colored</i>
	Contr./C-E default	<i>colored</i>
	Subtyped	<i>embedded, sanded, tinged, petrified</i>
A	Modified default	<i>generated, manufactured air-born</i>
	Contr./C-E default Subtyped	
T	Modified default	
	Contr./C-E default Subtyped	

Table 3: *Dust*

Qualia	PP Modifier	Examples
F/C	Modified default	<i>colored, scented, flavored, atomized, crystallized</i>
	Contr./C-E default	
	Subtyped	
A	Modified default	<i>produced, harvested, extracted melted</i>
	Contr./C-E default Subtyped	
T	Modified default	<i>used, utilized, ingested</i>
	Contr./C-E default Subtyped	

Table 4: *Water*

9 Conclusion

The goal of this paper has been to prove that the notion of informativeness (traditionally ascribed

Qualia	PP Modifier	Examples
F/C	Modified default	<i>colored, equipped, bodied, shaped</i>
	Contr./C-E default Subtyped	<i>winged, twin-engined, armed</i>
A	Modified default	<i>made, developed, produced, constructed, manufactured, created</i>
	Contr./C-E default Subtyped	<i>manufactured</i>
T	Modified default	<i>used, flown, operated, utilized</i>
	Contr./C-E default Subtyped	<i>used, utilized</i>

Table 5: *Aircraft*

Qualia	PP Modifier	Examples
F/C	Modified default	<i>colored, bodied</i>
	Contr./C-E default Subtyped	<i>aromatized</i>
A	Modified default	<i>made, produced, grown, created, farmed, harvested</i>
	Contr./C-E default Subtyped	<i>crafted</i>
T	Modified default	<i>drunk, consumed</i>
	Contr./C-E default Subtyped	

Table 6: *Wine*

to the pragmatic domain and not sufficiently formalized before in the literature) can be accounted for compositionally at the phrasal and clausal level, and that the degree of informativeness of a given expression can be calculated by combining the model generation strategy with some of the basic notions of GL: first and foremost, the values provided by the qualia structure, as well as the GL typology of arguments (including *default* and *shadowed*). We suggested that, for a construction to be acceptable, it must be *consistent* (realizable in at least some situations) and *informative* (not satisfied in at least some situations). The contribution of an argument to the construction is only informative if it does not refer to an inherent property of the syntactic head (be it a verb, as in predicative constructions, or a noun, as in modification constructions); in terms of *qualia informativeness*, it must not refer to default qualia values of the syntactic head. We also proposed that the degree of informativeness of a given construction is crucially determined by the compositional mechanism involved in its derivation, and ranked the type satisfaction mechanisms accordingly: introduction is the most informative one, and type matching and exploitation are zero informative. We showed that this approach is borne out by corpus data by examining naturally occurring PP-N combinations.

Ongoing research elaborates on the formal details of the mechanism outlined in this paper and extends its application to a wide range of linguistic phenomena whose properties are determined by the general informativeness requirement.

Acknowledgments

The contribution by Olga Batiukova was partially financed by the Ministry of Economy and Competitiveness of Spain under Grant No. FFI2012-33807.

References

- F. Ackerman and A.E. Goldberg. 1996. Constraints on adjectival past participles. In A.E. Goldberg, editor, *Conceptual Structure, Discourse and Language*, pages 17–30. CSLI Publications.
- N. Asher and J. Pustejovsky. 2013. A type composition logic for generative lexicon. In *Advances in Generative Lexicon Theory*, pages 39–66. Springer.
- P. Blackburn and J. Bos. 2008. Computational semantics. *THEORIA. An International Journal for Theory, History and Foundations of Science*, 18(1).
- J. Bresnan. 1995. Lexicality and argument structure.
- N. Chomsky. 1993. *Lectures on government and binding: The Pisa lectures*, volume 9. Walter de Gruyter.
- W. Croft. 2005. Logical and typological arguments for radical construction grammar. *Construction Grammars: Cognitive grounding and theoretical extensions*, pages 273–314.
- D. Dowty. 1979. *Word meaning and Montague grammar: The semantics of verbs and times in generative semantics and in Montague's PTQ*, volume 7. Springer.
- D. Embick. 2004. On the structure of resultative participles in english. *Linguistic Inquiry*, 35(3):355–392.
- A.E. Goldberg and F. Ackerman. 2001. The pragmatics of obligatory adjuncts. *Language*, 77(4):798–814.
- A. Goldberg. 1998. Semantic principles of predication. *Discourse and Cognition: Bridging the Gap, CSLI Publications, Stanford University, Stanford, CA*, pages 41–54.
- J. Grimshaw and S. Vikner. 1993. Obligatory adjuncts and the structure of events. In E. Reuland and W. Abraham, editors, *Knowledge and language*, volume II, pages 145–159. Kluwer Academic Publishers, Dordrecht.
- L. Horn. 1996. Presupposition and implicature. *The handbook of contemporary semantic theory*, pages 299–319.
- Y. Jung. 1997. Obligatory adjuncts. *Texas Linguistic Forum*, 38:161–171.
- A. Kilgarriff, P. Rychlý, P. Smrž, and D. Tugwell. 2004. The sketch engine. In *Proceedings of the Eleventh EURALEX International Congress. Lorient, France: Universite de Bretagne-Sud*, pages 105–116.
- K. Konrad. 2004. *Model generation for natural language interpretation and analysis*, volume 2953. Springer.
- R. Langacker. 1991. *Foundations of Cognitive Grammar*. Stanford University Press.
- S. Levinson. 2000. *Presumptive meanings: The theory of generalized conversational implicature*. The MIT Press.
- K. McRae and K. Matsuki. 2009. People use their knowledge of common events to understand language, and do so as quickly as possible. *Language and linguistics compass*, 3(6):1417–1429.
- J. Pustejovsky and O. Batiukova. forthcoming. Compositionality and information.
- J. Pustejovsky and E. Jezek. 2008. Semantic coercion in language: Beyond distributional analysis. *Italian Journal of Linguistics*, 20(1):175–208.
- J. Pustejovsky. 1995. *Generative Lexicon*. Cambridge (Mass.): MIT Press.
- J. Pustejovsky. 2011. Coercion in a general theory of argument selection. *Linguistics*, 49(6):1401–1431.
- J. Pustejovsky. 2013. Type theory and lexical decomposition. In *Advances in Generative Lexicon Theory*, pages 9–38. Springer.
- R. Van Valin. 1990. Semantic parameters of split intransitivity. *Language*, 66:221–260.

