Polysemy, Polyvalence, and Linking Mismatches: The Concept of Rain and its Codings in English, German, Italian, and Spanish*

(Polissemia, Polivalência e Falhas de Ligação:
O Conceito de Rain (chuva, chover) e suas Codificações
em Inglês, Alemão, Italiano e Espanhol.)

Dietmar Zaefferer
(University of Munich)

ABSTRACT: *Based on an ontological frame for comparative onomasiological lexicology which embeds the RAIN eventity type as an exponent-shaped activity into a general linguistic ontology as well as an ontology of motion, the words for RAIN in two Germanic and two Romance languages are compared with respect to their semantic and syntactic variability. It turns out that the Germanic forms are more flexible than the Romance forms. English rain has the highest variability: It can be used with eight valencies, a nominal and seven verbal ones, and with five different meanings, and it is the only one to display a valence that combines an expletive subject with a cognate object as well as a causativized eventity meaning. On the other hand only German regnen has a resultativized meaning. The Romance languages too show some internal variance: In allowing both an impersonal construction with a generalized meaning and a source construction, Spanish llover is slightly more flexible than Italian piovere. Finally, an Optimality Theoretic account of the data is adumbrated in terms of three principles: Avoid linking mismatches, Avoid heterogeneity of sentence patterns, and Avoid redundancy.*
KEY-WORDS: *Polysemy; Valency; Meaning; Linking pattern.*

RESUMO: *No âmbito de uma lexicologia comparativa onomasiológica, e com base em um esquema ontológico que define o tipo de eventividade associado a RAIN (chuva,*

chover) como uma atividade moldada por expoente, as palavras para RAIN em duas línguas germânicas e duas línguas românicas são comparadas com respeito à sua variação semântica e sintática. Observa-se que as formas germânicas são mais flexíveis que as formas românicas. O item lexical rain, do inglês, apresenta a maior variação. Ele pode ser usado com oito valências diferentes, uma nominal e sete verbais, e com cinco sentidos diferentes, sendo o único que apresenta uma valência que combina um sujeito expletivo com um objeto cognato, assim como um sentido de eventividade causativizada. Por outro lado, somente o item lexical regnen, do alemão, apresenta um sentido resultativo. As línguas românicas também apresentam alguma variação interna. O verbo espanhol llover, ao possibilitar uma construção impessoal com um sentido generalizado e uma construção indicativa de origem, é ligeiramente mais flexível que o verbo italiano piovere. Finalmente, é apresentada uma análise dos dados pela Teoria da Otimalidade, com base em três princípios: evite falhas de ligação, evite heterogeneidade de padrões sintáticos e evite redundância.

Palavras-chave: *Polissemia; Valência; Sentido; Padrões de ligação.*

0. Introduction

Polysemy, the association of a single phonological shape with two or more systematically interrelated meanings, mostly a whole bundle, is ubiquitous. That's a truism: Natural human languages are simply too efficient and too flexible to do without. Monosemy is the marked case, polysemy the unmarked one. But this holds only on the lexical level, at the level of word tokens outside of helpful context. On the sentential level, the default level of complete utterances, the opposition is exactly the other way round: Monosemy is the unmarked case, and polysemy, in this context called ambiguity, the marked one.

The reason for this lies in the close connection of lexical polysemy with polyvalence: The different meanings of a polysemous item tend to be triggered by different valency patterns, i.e. they often have different grammatical distributions, or, more precisely, they call for different kinds of completion in order to build whole phrases or utterances.

The notion of valence or valency comes in two variants: as grammatical valence and as semantic valence. If we follow Cruse (1986: 80) in calling a reading of a lexeme a lexical unit, then we can define in a very general

fashion the *grammatical valency* of lexical unit the set of operations that are required to form appropriate complete linguistic items from it. An appropriate complete linguistic item to be formed from a noun reading is for instance a corresponding noun phrase, and an appropriate complete linguistic item to be formed from a verb reading is a clause or a sentence.

Take for instance the Spanish verb *saltar* in its reading 'to hop'. In a situation where it is clear that we are speaking about a particular child, the only operation required to form a complete declarative senctence is the selection of an appropriate form from the paradigm of *saltar*, namely *Salta*. In a situation with several topical persons, a second operation is required that consists in the concatenation of a referring noun phrase: *El niño salta*. Note that in the first case a position is left open that is filled in the second case: The position of that element in the clause that plays the subject role in it. So it turns out that the grammatical valence of a lexical unit depends on pragmatic circumstances. Therefore the lexicon cannot indicate the actual grammatical valence of such a unit, it rather has to specify its *grammatical valence potential* together with the conditions that determine the choices from it.

The second form of valency is semantic valency. Again, we will start with a very general notion and call *semantic valency* of a lexical unit the set of operations that are required to form appropriate complete semantic items from it. An appropriate complete semantic item to be formed from a noun reading is for instance a concept of a restricted act of reference, and an appropriate complete semantic item to be formed from a verb reading is a concept of an illocutionary act or of a proposition.

Take again the Spanish verb *saltar* in its reading 'to hop'. In a situation where it is clear that we are speaking about a particular child, four operations are required to form the concept of a complete assertive illocution: First the creation of the concept of an act of referring to the the topic situation, second the creation of the concept of an act of referring to the topical child in this situation, third the creation of an instantiation of the concept of hopping in the topic situation, and fourth the creation an instantiation of the concept of the speaker claiming that in the topic situation the topical child fills the exponent role of the event token, namely the given case of hopping. Notions like exponent role (the role of the entity that displays the behavior in question) are called semantic roles.

Again, in a situation with different topical persons things change slightly: The same four conceptual operations are needed to form the concept of a complete assertive illocution with the exception that the second operation consists in the creation of the concept of an act of referring to one of the topical people in the topic situation that is restricted by the presupposition that it be the only or most salient child among them. Note that in both cases the fourth operation of the semantic valency of the abovementioned reading of *saltar* involves the idea that the exponent role of the given case of hopping is filled. So the grammatical asymmetry between the two sentences, one without and the other one with subject constituent, does not carry over to the conceptual content: in both cases the conceptual role of the exponent of hopping is filled. Thus unlike the grammatical valence the semantic valence of lexical unit does not depend on pragmatic circumstances and there is no need to distinguish between actual semantic valence and semantic valence potential.

This points to the important fact that although grammatical and semantic valency have the tendency to match, they do not have to match. How come? We have just seen that alongside with grammatical form and semantic content there is a third factor that plays a role in linguistic encoding: pragmatic circumstances and more generally presentation structure. Presentation structure (often also called information structure) is understood in the present paper as consisting of three dimensions of variation and correspondingly six pragmatic roles: (a) what is talked about (*Topic*) and what is said of this (*Predication*), (b) the input stage of the discourse, determining what is *Given* and what *New*, and (c) the intended output, distinguishing between *Focus* and *Background*. In other words, presentation structure means the incremental build-up of conceptual content in a given communication situation. One of the principles of presentation structure says that when something is obvious it needn't be encoded lest the grammar requires that. The abovementioned case of *salta* illustrates the well-known fact that in pro-drop languages like Spanish elements bearing semantic roles needn't be expressed when the grammatical role of this expression would be subject and when their pragmatic roles are Topic, Given, and Background.

In view of the rather encompassing notions of grammatical and semantical valency just outlined it will be helpful to introduce some further terms for the more narrow notions that are often associated with these expressions.

Let *a* be a lexical unit. We then call the *grammatical role inventory* of *a* those grammatical roles that according to *a*'s grammatical valency are compatible with *a*, and *core g-role inventory* of *a* those elements of the grammatical role inventory of *a* that must be overtly realized on grammatical grounds together with the pragmatic conditions under which this holds.

Let *a* be as above. We then further call the *semantic role inventory* of *a* those semantic roles that according to *a*'s semantic valency are compatible with *a*, and *core s-role inventory* of *a* those elements of the semantic role inventory of *a* that must be filled on analytical grounds (i.e. the role-giving concept would be unconceivable without them).

Now we are in a position to understand the notion of linking mismatches in the title of this paper. Linking is the association of grammatical roles on the one hand with semantic and pragmatic roles on the other. Linking has therefore two aspects: decoding (given an expression bearing a grammatical role in an utterance expression, what semantic and pragmatic role is borne by its denotation?) and encoding (given the semantic and pragmatic role of an element of an utterance meaning, what is the grammatical role of the expression that encodes it, if there is any?).

Linking mismatches occur when at least one aspect of linking is problematic, they can therefore subdivided in different types. *Overcoding mismatches* occur where at least one element of the grammatical role inventory of a lexical unit does not denote, i.e. cannot be interpreted as referring to some element of the semantic role inventory of that unit. This is also known as *expletion*, examples from English being *there* in presentational constructions and *it* in weather verbs.

Undercoding mismatches come in two varieties. *Weak undercoding mismatches* are found where at least one of the elements playing a semantic role from the core s-role inventory of a lexical unit is not expressed, i.e. denoted by some constituent that plays a role from the grammatical role inventory of that unit. This occurs quite often and is a well-studied phenomenon, at least with the unary use of transitive verbs as in *Mary is eating*, where logically there must be something she is eating, but this need not be specified as long as it is the default undergoer of eating, namely food. Pustejovsky (1995: 64) speaks here of default arguments.

Much less frequent and also less well understood are *strong undercoding mismatches*. They arise where one semantic role bearing element is not only not expressed, but in addition not even expressible. An example is the German intransitive verb *zubeissen* 'to bite', which cannot be constructed with an object or prepositional phrase that denotes the thing bitten, even though this is in a semantic role that is required on analytic grounds (there is no case of *zubeissen* without something being bitten).

The last question to be answered in this already somewhat lengthy introduction regards the relation between the topic of polysemy, which of course is a semasiological one, and the case study of the codings of the RAIN concept announced in the subtitle of this paper: How can an onomasiological investigation help in the study of a semasiological problem? The idea is to start from an interculturally and interlinguistically stable and rather well-defined concept and to ask for those lexemes that code this concept in different languages: Are they monosemous or are they polysemous? If the latter is the case, to what extent do their reading patterns coincide and to what extent do they differ? And finally: Is it possible to explain the findings at least partially?

1. Talmy's Law

The history of modern decompositional lexical typology goes back to 1985 when Leonard Talmy's groundbreaking article Lexicalization patterns: Semantic structure in lexical forms (Talmy 1985) appeared. Comparing the lexical coding of motion events in a sample of typologically quite heterogeneous languages Talmy found that alongside with the superconcept of motion itself four additional factors are of prime importance: The entity that is in motion (its exponent, sometimes also called theme), the background against which motion takes place, the path along which motion takes place, and the manner or cause of motion. Here are the factors together with their labels:

0.	The superconcept	motion	S
1.	The exponent of motion	exponent	E
2.	The background of motion	ground	G
3.	The path of motion	path	P
4.	The manner or cause of motion	manner/cause	M

What he discovered is a very remarkable statistical universal of lexical typology that I would like to call Talmy's Law:

(T) Natural human languages tend to co-encode with the concept of motion exactly one additional factor of motion.

(T) entails among other things that languages where the shortest translation of *walk* or *enter* would correspond to something like 'move on ones own feet' or 'move into' are equally dispreferred as are languages with monomorphemic verbs that translate as 'roll downwards over grass as or like a rock'. *Walk* on the other hand is well-behaved with respect to (T) since it co-encodes with the superconcept only manner of motion ([S + M]), leaving the coding of path information to other components of the clause (*walk in* [[S + M] & P]), and the same holds for the encoding of the exponent and the background of motion. *Enter* conforms with (T) in a different way: It co-encodes with the motion its path ([S + P]), leaving to other elements the coding of manner (*enter walking* [[S + P] & M]) or other information.

Talmy's typological distinction between between what he calls verb-framed and satellite-framed languages like Spanish on the one hand and English on the other, the former preferring path verbs in the coding of motion events whereas the latter prefer manner verbs, has spurred a large discussion (cf. e. g. Slobin 2000 for a modified Whorfian interpretation of this opposition and Papafragou et al. 2001 for an opposing view and pertinent data) which will not be commented upon here since this paper is concerned mainly with polysemy. But it will be useful to elaborate a little on Talmy's analysis since our case study involves motion events of a different kind.

2. An ontological frame for comparative lexicology

2.1. *General linguistic ontology*

Before we get down to a notional analysis of the concept of RAIN, it will be useful to provide a little background in terms of a general linguistic ontology, where linguistic ontology means a conceptualization or

categorization of what normal everyday human language can talk about, not the terminologically enriched special languages of specialists. The most basic distinction is that between concrete and abstract entities, and since this disctinction is based upon spatiotemporal location, a third basic category has to be considered, namely spatiotemporal enties. They include of course local and temporal spaces, but directions are also needed and, especially for motion, inherently directed and bounded spaces, namely paths.

Boundedness is also a prominent property in the ontology of concrete entities. One of the most important categories in this domain is that of a world segment or situation, since as we have seen in the case of the hopping child from the introduction, people in general do not talk about the whole universe but rather about rather small portions of it, which are called situations. Situations are conceptualized as containers and therefore as inherently bounded, although their boundaries may coincide in the limit case with those of the universe.

Situations are inhabited by individuals, but also by entities like substances or collections that are not inherently bounded and that can be further divided without losing their integrity. Substances and collections together will therefore be called dividuals, and dividuals together with individuals will be called inventities, since they build the inventory of situations and they are the protagonists of instances of the next ontological category, eventities. Eventities, a shorthand for events and similar entities, include the Aristotle – Kenny – Vendler categories of states, activities, achievements, and accomplishments, but also properties, semelfactives and intergressives, in short, the eventity ontology presented here is a taxonomy of aspectualities.

Finally, abstract entities are all those that are neither spatiotemporal entities nor spatiotemporally located. They include propositions and speech act types, but not speech act tokens. Table 1 below shows the structure of the general linguistic ontology that serves as a background to the following investigations.

I. Spatiotemporal entities

    A. Local spaces   (not inherently directed)

    B. Directions   (not inherently bounded)

    C. Temporal spaces   (inherently directed)

    D. Paths   (inherently directed and bounded spaces)

        1. goal bounded paths   (partially bounded)

        2. source bounded paths   (partially bounded)

        3. doubly bounded paths   (completely bounded)

II. Concrete entities   (spatiotemporally located)

    A. Situations   (spatiotemporally coherent containers; inherently bounded)

    B. Inventities   (spatiotemporally coherent contents; may include roles for other entities)

        1. Individuals   (inherently bounded)

            a. Absolute individuals

            b. Relational individuals   (presuppose the existence of a different entity that fills a role in them)

        2. Dividuals   (not inherently bounded)

            a. Substances   (non-atomic)

            b. Collections   (atomic)

    C. Eventities   (spatiotemporally coherent contents; tend to include roles for other entities)

        1. Characteristics   (not inherently bounded)

            a. Properties   (inalienable)

            b. Stages   (alienable)

               i. Static: States   (not intensity-gradable)

               ii. Dynamic: Activities   (intensity-gradable)

        2. Events   (inherently doubly bounded; not extended)

            a. Non-transitional: Semelfactives

            b. Transitional: Achievements

        3. Processes   (inherently doubly bounded; extended)

            a. Non-transitional: Intergressives

            b. Transitional: Accomplishments

III. Abstract entities   (not spatiotemporally located: qualitative and quantitative concepts, conceptual systems, propositions etc.)

Table 1. Linguistic Ontology

2.2. Linguistic ontology of motion

Motions are of course eventities and they can occur in different aspectualities. If they are not inherently bounded, they are alienable and therefore stages, and they are dynamic and therefore activities. If they are inherently bounded, they can be transitional, i.e. conceptualized in terms of a change, such as *travel to London* (extended, hence process) or *arrive at the airport* (not extended, hence event), or not transitional, e.g. *twitch once* or *circle the bloc*k.

But there are other distinctions among motions that are not implied by the general ontololgy of the last section. One of them becomes immediately visible when we compare the codings of the superconcept itself in the four sample languages:

	intransitive verbs	transitive verbs	nouns
English:	*move*	*move*	*motion, movement, move*
German:	*sich bewegen*	*bewegen*	*Bewegung*
Spanish:	*moverse*	*mover*	*movimiento*
Italian:	*muoversi*	*muovere*	*movimento, moto*

It turns out that English is the only language in the sample that uses the same verb for the coding of the motion itself and for the coding of the causation of some motion. We will speak of causedness value of motion, which may be either causeless (the cause may be there, but it is not co-conceptualized) or caused. Note that English does code the distinction, but through the valencies of a polysemous verb: One reading is associated with the intransitive valency, the other one with the transitive valency.

Caused motion means that motion is conceptualized as the effect in a cause-effect pair. If motion is conceptualized as the cause in a cause-effect pair, we speak of effecting motion. So looking to the right, so to speak, we can make a second binary distinction: effectless motion (the effect may be there, but it is not co-conceptualized) versus effecting motion. The effect may be motion in turn, but it need not. The polysemous verb *sweep* (cf. Rapaport Hovav/Levin 1995) illustrates all three cases:

(1)   A huge bird swept down.

(2)   Phil swept the floor.

(3)   Phil swept the crumbs into the corner.

The intransitive use in (1) codes an effectless motion, the transitive uses in (2) and (3) code effecting motions: In (2) the effect is not a motion, rather a state of being clean, whereas in (3) the effect of the sweeping motion is the caused motion of the crumbs into the corner.

The examples show a third dimension of variation that is important with motions. If the bird sweeps down then the whole exponent is moving along a path, so we speak of a path motion. If Phil sweeps the floor and by the same token the crumbs into the corner, then Phil as whole is not necessarily moving along a path, he may just move his arms and a broom. So it makes sense to distinguish between path motion and internal motion, depending on the exponent: If Phil lifts his arm, the arm makes a path movement and Phil an internal movement. If Phil walks on the spot he makes an internal motion, if walks across the lawn he makes a path motion.

These are three binary dimensions of variation. The remaining dimensions are *n*-ary: Manner of motion, as mentioned above, and exponent of motion. Taking stock we can state that on a priori grounds every motion must have a value in at least the following five dimensions:

- causedness value of motion: causeless motion   /   caused motion
- effectivity value of motion: effectless motion   /   effecting motion
- locality value of motion: internal motion   /   path motion
- manner of motion: walking, jumping, swimming …; slow, fast, …
- exponent of motion   individuals: people, animals, rocks…; collections, substances

A little more formally, we can formulate the following ontological postulates (i_n, e_n, and l_n are parameters that range over inventities, eventities, and spatiotemporal entities, respectively; the colon codes the instantiation relation such that 'x : Y' may be read as 'x is a token of type Y', and 'x [x : Y]' codes existential quantification 'there is an x such that x is of type Y')

If in some situation s_0 there is a motion eventity e_0, then it must be the case that in s_0

(a) i_0 [i_0 : exponent(e_0)]

(b) e_1 [e_1 : manner(e_0)]

and it may be the case that in s_0

(c) l_0 [l_0 : path (e_0)]

(d) e_2 [e_2 : effect(e_0)]

(e) e_3 [e_3 : cause(e_0)]

If e_0 is a path movement of i_0 in s_0 and l_0 is its path, then it must be the case that in s_0

(f) l_1 [l_1 : direction(l_0)]

and it may be the case that in s_0

(g) l_1 [l_1 : goal(l_0)]

(h) l_2 [l_2 : source(l_0)]

(i) l_3 [l_3 : reference point(l_0)]

(j) l_4 [l_4 : background(l_0)]

Finally, if i_0 is the exponent of an eventity e_0 in s_0, then it may be the case that in s_0

(k) i_1 [i_1 : instrument(i_0, e_0)]

i.e. there is an instrument used by the exponent in displaying the eventity.

2.3. Onomasiological lexicology

According to what we have called Talmy's Law, monomorphemic codings of motion eventities tend to co-encode exactly one additional factor with the concept of motion itself. Let us say that in such a case this conceptualization of a motion eventity is shaped by the factor in question, be it an obligatory or an optional one. Codings of such conceptualizations may be transparent or opaque with respect to this shaping. Here are a few examples:

	example	transparency
cause-shaped caused motion		
instrument-shaped		
caused motion	*hand*, *shovel*	transparent
cause-shaped effecting motion		
instrument-shaped		
effecting motion	*sweep / scopare*	opaque / transparent[1]
path-shaped motion		
direction-shaped motion	*rise*	opaque
goal-shaped motion	*enter*	opaque
source-shaped motion	*exit*	opaque
reference point-shaped motion	*pass*	opaque
manner-shaped motion	*crawl*	opaque
exponent-shaped motion	*rain*	transparent

It turns out the the concept of RAIN as coded by the English verb *to rain* and its counterparts in many languages is a transparent exponent-shaped concept paraphrasable as 'to do what the inventity rain prototypically does'.

---

[1] *scopa* is the Italian word for broom, so *scopare* transparently codes that a broom is the instrument of the coded activity, whereas *sweep* hides the fact that sweeping is prototypically done with a broom or brush.

3. The Concept of RAIN and its Codings

3.1. *The decompositional conceptual structure of RAIN eventities*

3.1.0. Mutual implication of eventity and inventity

We have analysed the eventity type RAIN as an exponent-shaped motion concept since its specificity is due to the inventity that normally displays this kind of motion, a large amount of about evenly distributed drops of water, in terms of the basic ontology outlined above a collection. The American Heritage Dictionary (fourth edition) gives as first reading of the intransitive verb *rain*: "To fall in drops of water from the clouds". The first reading of the corresoponding noun, according to the same dictionary, is: "Water condensed from atmospheric vapor and falling in drops". After some reformulation the relation between the verbal and the nominal concept becomes more transparent: 'Drops of water falling from the clouds' and 'To fall in drops of water from the clouds'. The exponent of the eventity is defined not only as a collection of quantities of a certain substance, but also as being involved into a falling motion. And the eventity is defined not only as a source-bounded path motion, but also as being displayed by a collection of drops of water.

So it is to be expected that the noun does not really need a substantial predicate, and the verb does not really need a substantial subject to make a proposition and that both can be used more or less interchangably. (4) and (5) demonstrate that this expectation is borne out at least for English:

(4)   It will rain tomorrow.

(5)   There will be rain tomorrow.

So we are entitled to suppose that the English syllable *rain* in such contexts shows polyvalence without polysemy since the category difference between noun and verb is of course also a difference in grammatical valency. Note that the noun, but not the verb has two slightly different readings, one being the eventity and this is shared with the verb, the other one being the inventity that prototypically is the exponent of such eventities, and this is not shared by the verb. (6) illustrates the former, (7) the latter:

(6)   The rain lasted for several hours.

(7)   The rain was very warm and contaminated with volcano ashes.

But these readings are conceptually so closely related that people don't feel a zeugma effect in cases like (8):

(8)   The rain was very warm and lasted for several hours.

This close and systematic relation is obviously a good candidate for what Pustejovsky (2002) calls a dot object. His example is the relation that holds between the two senses of *book*, as a physical object and as a certain amount of information, which is a container-content relation. The senses of the *rain* noun are related by a different dot object, the motion-exponent relation and the same holds for the inventity reading of the noun and the basic eventity reading of the verb.

3.1.1. Underlying inventity type: type and direction of displayed eventity are implied

Now we are in a position to give a formal characterization of the decompositional conceptual structure of the concept that constitutes the inventity reading of the noun *rain* and its counterparts in other languages. The notation is as above, with the following additional conventions: '[x : Y]' stands for the type of those x that are of type Y, 'x [x : Y] : Z' for the condition that all tokens of type Y are also of type Z, '+' joins types and '&' joins conditions.

So the first concept of RAIN to be defined here is the following:

RAIN^1 (basic inventity) : =

[i_0 : collection + distributed & i_1 [i_1 : member(i_0)] : unit(water) &

$\quad$ e_0 [i_0 : exponent(e_0) & e_0 : path-motion + activity &

$\quad\quad$ l_0 [l_0 : path(e_0) &

$\quad\quad\quad$ l_1 [l_1 : direction(l_0) + downwards] &

$\quad\quad\quad$ l_2 [l_2 : source(l_0) + location(clouds)]]]]

This says that an entity x is a case of RAIN[1] just in case x is an inventity that is a collection of distributed objects such that each member of x is a drop, or more generally unit, of water and there is an eventity displayed by x which is a path-motion and an activity such that the source of the path of this motion is the location of some clouds and the direction of this path is downwards. This concept can be coded only by a noun in all four sample languages and the codings are given in the order English, German, Spanish, Italian in (9) – (12):

(9)   rain

(10) Regen

(11) lluvia

(12) pioggia

3.1.2. Basic eventity type: Type of exponent inventity and direction are implied

Given the fact that according to our observation in the concept of RAIN inventity and eventity are like different sides of the same coin, it is easy to derive a formal definition of the basic eventity type from the definition above:

RAIN[2] (basic eventity) : =

$[e_0$: path-motion + activity &

$\quad i_0$ $[i_0$: exponent(e_0) + collection + distributed & $\; i_1$ $[i_1$: member(i_0)] : unit(water)] &

$\quad l_0$ $[l_0$: path(e_0) &

$\qquad l_1$ $[l_1$: direction(l_0) + downwards] &

$\qquad l_2$ $[l_2$: source(l_0) + location(clouds)]]]

The nominal codings of this concept in the four sample languages are the same as above ((9) – (12)), but now there are also verbal codings and the examples for these are embedded in full sentences:

(13) It is raining.

(14) Es regnet.

(15) Llueve.

(16) Piove.

The comparison shows that in the Romance expressions the impersonal verb suffices to constitute a clause (a pronominal subject would be ungrammatical), whereas the Germanic examples would be incomplete without the expletive subject. Thus we can state for (13) and (14) a double linking mismatch, consisting of an overcoding (non-denoting subject) and a weak undercoding (unexpressed participant), whereas (15) and (16) display only the latter.

Schematically, the linkings between clause structure and eventity structure can be represented as follows (LP stands for linking pattern, GR for grammatical relation, SC for semantic constituent):

(LP 1)	English, German	GR	1. subject	2. Ø	3. predicate
		SC	1. Ø	2. exponent	3. eventity
(LP 2)	Spanish, Italian	GR	1. Ø	2. predicate	
		SC	1. exponent	2. eventity	

But this picture implies a slight oversimplification. As we have seen in the conceptual decomposition, the exponent of the eventity is not really unexpressed, it is co-encoded in the verb and so it would be simply redundant to express it: (17) is at least very strange whereas (18) is fine.

(17) ?Piove una pioggia.

(18) Piove una pioggia lenta e leggera.   'It is raining a slow and light rain.'

This indicates that we are faced with a case of what Pustejovsky (1995: 65) calls a shadow argument. Shadow arguments share with default arguments the property that they needn't be expressed by a separate phrase, but unlike default arguments, they even mustn't be expressed except the expression is not redundant or some other special condition obtains. When

they are expressed, traditional grammar speaks of internal or cognate objects.

Note that here we have a new valency frame without a new meaning, since the type of RAIN[2] given above will only be enriched by the addition of a specified direct object. And it looks like all our sample languages offer this option with the single exception of German (examples are taken from the internet):

(19) It is raining a hard rain.

(20) ?Es regnet einen heftigen Regen.   'It is raining a strong rain.'

(21) Piove una pioggia quieta e persistente.   'It is raining a calm and persistent rain.'

(22) Llueve una lluvia soleada, brillante…..   'It is raining a sun trenched,brilliant rain.'

But is this analysis of the Romance data correct? Are the verb forms in (21) and (22) really impersonal forms? Or are they normal agreeing forms? Then they wouldn't be subjectless sentences anymore and the agreement would show in plural forms. Example (23)[2] supports this view.

(23) Este año han llovido lluvias tan copiosas, que ...
       'This year so copious rains have rained that ...'

So the Romance cases are better analysed as cases of internal or cognate subjects[3] leaving English as the only language in the sample with a transitive valency frame involving an expletive subject and a cognate object.

(LP 3)   English   GR   1. subject   2. cognate object   3. predicate
                       SC   1. Ø   2. exponent   3. eventity

On the other hand, the cognate subject option (mentioned already in Quirk et al. 1985, ch. 10.29 with the example *The frost froze hard*) is of course not only open to Romance languages, witness (24) and (25):

---

[2] From: "Historia general de las relaciones exteriores de la República Argentina", Obra dirigida por Andrés Cisneros y Carlos Escudé (http://www.argentina-rree.com/4/4-012.htm)

[3] This implies that the above translations are incorrect and have to be replaced by 'A calm and persistent rain is raining.' for (21) and 'A sun trenched,brilliant rain is raining.' for (22).

(24) The rain is raining all around. (Robert Louis Stevenson[4])

(25) Denn der Regen, der regnet jeglichen Tag. (Hermann Salomon Mosenthal[5])
'For the rain, it rains each and every day.'

These examples show by the way that the rule requiring redundant, unmodified shadow arguments to be left unexpressed has exceptions in poetry. Here is the linking pattern for the cognate subject construction:

(LP 4)	English, German, Spanish, Italian	GR	1. cognate subject	2. predicate
		SC	1. exponent	2. eventity

So far, the RAIN verbs have been shown to be polyvalent (in different languages to different degrees), but not polysemous. The following examples will change that picture.

3.1.3. Weakened eventity type: Type of exponent inventity not implied, direction implied

Cognate arguments are in general only admissible when they strengthen the underlying concept. Other, incompatible arguments should be inadmissible since they would make the resulting concept contradictory. However, examples like (26) etc. are quite normal.

(26) It rained confetti as the Big Apple prepared for the big night.

(27) Es regnete Konfetti.

(28) Llovió confeti.

(29) *È piovuto coriandoli.

The reason is that the underlying concept is first weakened as far as necessary to make the argument compatible, and then accordingly

---

[4] From his collection 'A Child's Garden of Verses and Underwoods' (1913).

[5] From the lyrics of an aria from 'Die lustigen Weiber von Windsor' by Otto Nicolai, entitled 'Als Büblein klein an der Mutterbrust'.

strengthened. Here is the decompositional structure of this weakened or generalized reading of the polysemous impersonal RAIN verbs in English, German, and Spanish:

RAIN[3] (generalized eventity) : =

$\{e_0$: path-motion + activity &

$\quad i_0$ $[i_0$: exponent(e_0) + collection + distributed & i_1 $[i_1$: member$(i_0)]$: unit$(X)]$ &

$\quad l_0$ $[l_0$: path(e_0) &

$\qquad l_1$ $[l_1$: direction(l_0) + downwards $]]\}$

With X = confetti this gives exactly the desired concept. There are two differences between this and the basic eventity type: The exponent still has to be a collection of distributed objects, but its member don't have to be units of water anymore, and the path of the motion still has to be downwards, but it doesn't have to originate in the clouds. The three languages in our sample that admit of this construction show two linking patterns:

(LP 5)	English, Geman	GR	1. subject	2. object	3. predicate
		SC	1. Ø	2. exponent	3. eventity

(LP 6)	Spanish	GR	1. object	2. predicate	
		SC	1. exponent	2. eventity	

If Italian (at least according to my informants) does not have this construction, this does not mean that *piovere* is unable to express this meaning. Here is a slightly different construction, available for all four languages, where exactly the same lexical meaning is involved as above, but with a different valency:

(30) ... where confetti rained at the end of the game ...

(31) Es regneten Konfetti.

(32) Del primero y segundo piso llovían confeti y luces de Bengala.
     'Confetti and Bengal lights rained from the first and second floor.'

(33) Sono piovuti coriandoli.

The linking patterns are the same for all sample languages:

(LP 7) English, German,   GR   1. subject   2. predicate
        Spanish, Italian   SC   1. exponent   2. eventity

Note that the German (31), although beginning with an expletive *es*, is not an expletive subject construction, since the subject appears in a postverbal position, so the preverbal slot is filled by a position expletive that is needed only in order to indicate verb second and that disappears as soon as the subject is fronted, cf. (34):[6]

(34) Konfetti regneten (*es) herab.

If it is correct that the linking patterns (LP 5) through (LP 7) all share the same lexical meaning, then any perceived differerences should be induced by different construction meanings. And so it seems to be, at least in German: Completely new exponents of RAIN eventities can be equally well realized as objects and as subjects, but with exponents that are anchored to given entities and hence not completely new, the latter option with a preverbal subject is preferred. (35) is from a German newspaper, (36) is constructed, the acceptablity difference is, however, a rather subtle one:

(35) Sternschnuppen aus diesem Meteoritenschwarm regnen alljährlich im
      November nieder.
      'Shooting stars from this meteorite cluster rain down every year in
      November.'

(36) ?Alljährlich im November regnet es Sternschnuppen aus diesem
      Meteoritenschwarm.

---

[6] The confusion of the subject expletive *es*, that belongs to the valency of some impersonal verbs, with the position expletive *es*, that disappears as soon as some other element fills the preverbal slot, doesn't seem to be a rare phenomenon, cf. e.g. Legendre 2001: 1 f.

3.1.4. Strengthened resultativized eventity type: Basic rain eventity conceived as effecting a path motion with overtly expressed exponent and path

It is well known that some intransitive and transitive constructions can be transitivized or retransitivized when they are accompanied by a resultative secondary predicate.

(37) The audience shouted the singer off the stage.

If *shout* is intransitively constructed, it admits no objects at all, if it is transitively constructed, it requires shoutable objects like commands or insults. In (37), however, it is constructed with a human undergoer which is understood as the exponent of a motion that is only indirectly expressed by the path phrase *off the stage*. This should, in principle, also be possible with impersonally constructed verbs, but out of our four sample languages, only German seems to realize this option.[7]

(38) Es regnete die Blüten vom Baum.

(39) *It rained the blossoms off the tree.

(40) *Llovió las flores del arbol.

(41) *È piovuto i fiori dal albero.

As usual with secondary predications a paraphrase is possible that joins two primary predications: It rained and as a result of that the blossoms came off the tree. So we have to add another valency and another meaning to the lexical entry for the polymorphous German verb *regnen*. The decompositional structure of this strengthened concept of RAIN has to provide space for a second eventity, the effected one, and for a second participant, the exponent of the effected path motion. It looks as follows:

---

[7] I am grateful to Steven Wechsler for correcting my mistaken assumption that English offers this option as well on the occasion of the Konstanz Workshop on Verbs, Arguments, and Polysemy (July 23-25, 2001).

RAIN[4] (resulativized eventity) : =

[e_0 : path-motion + activity &

$\quad i_0 \quad$ [i_0 : exponent(e_0) + collection + distributed & i_1 [i_1 : member(i_0)] :
$\qquad$ unit(water)] &

$\quad l_0 \quad$ [l_0 : path(e_0) &

$\qquad l_1 \quad$ [l_1 : direction(l_0) + downwards] &

$\qquad l_2 \quad$ [l_2 : source(l_0) + location(clouds)]] &

$\quad e_2 \quad$ [e_2 : effect(e_0) + path-motion &

$\qquad i_1 \quad$ [i_1 : exponent(e_2) + X] &

$\qquad l_3 \quad$ [l_3 : path(e_2) + Y]]]

Substituting 'the dust' for X and 'off the roof' for Y models exactly the semantic operations that correspond to adding a direct object and a directional phrase. Since two subevents are involved, the corresponding linking pattern looks a little more complicated:

(LP 8) German	GR	1. subject	2. Ø	3. object	4. Ø	5. directional	6. predicate
	SC	1. Ø	2. c-exponent	3. e-exponent	4. e-eventity	5. e-path	6. c-eventity

This shows, firstly, the same double linking mismatch as (LP1) since the subject does not denote and the exponent of the cause-event (the c-exponent) is not overtly expressed, but second, it shows another undercoding mismatch in that the effect-eventity (the e-eventity) is not overtly expressed either and has to be inferred from its exponent and its path.

From this it is easy to predict that a counterpart to (LP4), the cognate subject construction, should be available for the resultative alternation as well, and so it is:

(42) Ein heftiger Regen hat alle Blüten vom Baum geregnet.
$\qquad$ 'A hard rain has rained all the blossoms from the tree.'

The corresponding linking pattern shows only one undercoding mismatch:

(LP 9) German GR 1. cognate subject 2. object   3.Ø   4. directional 5. predicate

                SC 1. c-exponent   2. e-exponent 3. e-eventity 4. e-path   5. c-eventity

3.1.5. Causativized eventity type: Cause is unspecified activity with overtly expressed agent, effect is weakened rain eventity with overtly expressed exponent and path

The last derived interpretation of the RAIN-verbs to be discussed here is in a way the counterpart of the resultative construction just discussed, where rain was the cause of some other eventity, a path motion. Here the rain eventity is the effect of some other eventity, which remains unspecified except for its agent. The clearest examples come from English:

(43) The mob rained stones on the police

(44) Congress rained money upon itself

(45) Skinner rained abuse down on Mulder's head

The formal meaning representation can be obtained in an almost completely compositional way from $RAIN^3$, the generalized eventity type, by adding as a cause for the rain eventity a second activity whose agent has to be specified. The specification 'almost' is necessary because there is one little difference: $RAIN^3$ doesn't require the overt specification of a goal for the path motion, $RAIN^5$ does.

$RAIN^5$ (causativized eventity) : =

$[e_0$: path-motion + activity &

    i_0 $[i_0 :$ exponent(e_0) + collection + distributed & $i_1 [i_1 :$ member$(i_0)]$: unit(X)] &

    l_0 $[l_0 :$ path(e_0) &

        l_1 $[l_1 :$ direction(l_0) + downwards] &

        l_2 $[l_2 :$ goal(l_0) + Y]] &

    e_2 $[e_2 :$ cause(e_0) + activity &

        i_1 $[i_1 :$ exponent(e_2) + Z]]]

Type representations for the examples above are obtained from RAIN[5] throught the following substitutions:

(43') X = stones,   Y = location(the police),   Z = the mob

(44') X = money,   Y = location(Congress),   Z = Congress

(45') X = abuse,   Y = location(Mulder's head),   Z = Skinner

The linking pattern shows the strong undercoding mismatch with respect to the causing activity.

(LP 10)	English	GR	1. subject	2.Ø	3. object	4. directional	5. predicate
		SC	1. c-exponent	2. c-eventity	3. e-exponent	4. e-goal	5. e-eventity

The other languages in our sample don't offer this option, they require at least a formal expression of the causing activity in term of a causative modal or light verb.

(46) a.   *Der Mob regnete Steine auf die Polizei.
     b.   Der Mob ließ Steine auf die Polizei regnen.

(47) a.   *El populacho llovió piedras sobre la policía.
     b.   El populacho hizo llover piedras sobre la policía.

(48) a.   *Il popolaccio piove pietre sulla polizia.
     b.   Il popolaccio fa piovere pietre sulla polizia.

But there seem to be exceptions to this rule. In the following examples no causative verb is used:

(49) Hier regnen die Wolken vier Dichter auf das Schiff herab.[8]
     'Here the clouds rain four poets down on the ship.'

(50) El Señor llovió grandes piedras sobre ellos.
     'The Lord rained big stones on them.'

---

[8] This example is taken with thanks from the Projekt Deutscher Wortschatz (http://wortschatz.uni-leipzig.de), which attributes it to the Kindler Literaturlexikon.

(50) is of course doubly special in that it involves an Almighty agent and biblical language, but (49) seems to be worth a closer inspection. Why is no causative verb used here although German in general seems to require one? In order to find an answer it is useful to remember that causativized RAIN eventities are analysed as consisting of two subeventities, the (unspecified) cause and the rain effect. For example the mob does something (presumably throwing) that has the effect that stones rain on the police. But what kind of activity of the clouds could have the effect that four poets rain down on the ship?

Before we answer this question it could be helpful to look at counterparts of this construction in the other languages. We already know that English doesn't need a causative verb, but with clouds as subject the situation is a little different: (51) illustrates that in this case the otherwise obligatory goal directional becomes optional. (52), the title of a song, shows that with this special subject Spanish requires neither a goal directional nor a causative verb.

(51) Clouds rained blood.

(52) Si las nubes llueven vino   'If the clouds rain wine'

All this makes it rather plausible to assume that (49), (51), and (52) do not involve two subeventities, the causing one and the resulting rain, but rather one. So they do not instantiate the causativized eventity concept RAIN^5, but rather a further concept that can be derived from the basic eventity concept RAIN^2 by promoting the source location to a source inventity, turning thus the eventity with only one participant into one that involves two participants, an agent and an exponent of the movement. This is not very surprising, since the conceptual closeness of sources and agents is rather well-known. Here are decompositional structure and linking pattern:

RAIN^6 (source agentive eventity) : =

$[e_0$: path-motion + activity &

$\quad i_0$ $[i_0$: exponent(e_0) + collection + distributed & i_1 $[i_1$: member(i_0)] : unit(X)] &

$\quad l_0$ $[l_0$: path(e_0) &

$\quad\quad l_1$ $[l_1$: direction(l_0) + downwards]] &

$\quad i_1$ $[i_1$: agent(e_0) + cloud]]

(LP 11) English, German, Spanish GR 1. source   2. object   3. predicate
                                         subject
                                   SC 1. agent   2. exponent   3. eventity

4. Conclusion

In this article, the codings of the concept of RAIN in four languages have been compared. It turned out that all of these codings are polymorphic, i.e. polysemous and polyvalent, but in various degrees. Six meanings were distinguished and twelve valency frames were found to be compatible with these meanings. Seven out of the eleven verbal valencies and corresponding linking patterns showed linking mismatches, four of them overcoding mismatches (expletion), five of them undercoding mismatches (expletion), and hence two of them double mismatches. The overcoding mismatches are of course found only in the two Germanic languages since this kind of expletion seems to be absent from the Romance languages. But apart from that, the different patterns did not coincide with the family boundaries. Table 2 below gives an overview of the findings.

It turns out that the English form is the most flexible one: It can be used with eight valencies, a nominal and seven verbal ones, and with five different meanings, three with the nominal valency, four with the verbal one, two with both. German is a little less flexible in that it has two different, albeit very closely related, forms for the nominal valency with its three meanings and for the seven verbal valencies with their four meanings. Note also that although the number of valencies and meanings of the verbal variant are the same in English and in German, the kinds are different: On the one hand, only German *regnen* has the resultativized meaning with its two valencies, on the other hand, only English *rain* has the expletive subject with cognate object valency and the causativized eventity meaning with its valency.

The Romance languages also show a slightly heterogeneous behavior: In allowing an impersonal construction with the generalized meaning and a source construction, Spanish *llover* is a little more flexible than Italian *piovere*.

	English	German	Spanish	Italian
M 1: RAIN[1] (basic inventity)				
V 1: noun	*rain*	*Regen*	*lluvia*	*pioggia*
M 2: RAIN[2] (basic eventity)				
V 1: noun	*rain*	*Regen*	*lluvia*	*pioggia*
V 2: verb with LP 1	*rain*	*regnen*	-	-
V 3: verb with LP 2	-	-	*llover*	*piovere*
V 4: verb with LP 3	*rain*	-	-	-
V 5: verb with LP 4	*rain*	*regnen*	*llover*	*piovere*
M 3: RAIN[3] (generalized eventity)				
V 1: noun	*rain*	*Regen*	*lluvia*	*pioggia*
V 6: verb with LP 5	*rain*	*regnen*	-	-
V 7: verb with LP 6	-	-	*llover*	-
V 8: verb with LP 7	*rain*	*regnen*	*llover*	*piovere*
M 4: RAIN[4] (resulativized eventity)				
V 9: verb with LP 8	-	*regnen*	-	-
V 10: verb with LP 9	-	*regnen*	-	-
M 5: RAIN[5] (causativized eventity)				
V 11: verb with LP 10	*rain*	-	-	-
M 6: RAIN[6] (source agentive eventity)				
V 12: verb with LP 11	*rain*	*regnen*	*llover*	-
Total number of valencies	8 (1 / 7)	1 / 7	1 / 5	1 / 3
Total number of meanings	5 (3 / 4)	3 / 4	3 / 2	3 / 2

**Table 2. Meanings, valencies, and linking patterns
of RAIN coding lexical units**

At the end of section 1, we have asked three questions: Are the different codings of the RAIN concept monosemous or are they polysemous? If the latter is the case, to what extent do their reading patterns coincide and to what extent do they differ? And finally: Is it possible to explain the findings at least partially? We are now in a position to answer the first two questions: All the codings are polysemous and the reading patterns coincide in differ in various degrees, that can be read off table 2. The hardest question is of course the last one. There are attempts in Optimality-Theoretic Syntax to account for the disjointness of subject pro-drop and subject expletion (cf. Grimshaw / Samek-Lodovici 1998), which shows up in the present

investigation as well. In the very same spirit one can speculate that there is a trade-off between a certain kind of iconicity (avoid linking mismatches) and a certain kind of uniformity (avoid heterogeneity of sentence patterns), and that the Romance languages rank the former over the latter, whereas the Germanic languages choose the opposite ranking. In the context of the present discussion a further principle comes into play: Avoid redundancy. When eventity and exponent inventity are interdefinable in the prototypical case, as is the case with the RAIN concept, then the solution to both problems, the mismatch and the pattern heterogeneity, namely the use of a cognate subject, is problematic itself because of its striking redundancy. So the patterning of data we have observed can be understood on the light of the interaction of these three forces.

But the more subtle differences among the two expletion languages as well as among the two pro-drop languages are still awaiting an explanation. At least, the differences have been stated here in a very concise and clear manner. And that seems to be the very prerequisite for any explanatory account. In a way, the present paper can be read as an implicit plea for a principled, systematic, precise and detailed presentation of the data in discussions related to questions of polysemy, polyvalency and linking.

REFERENCES

BARBOSA, P., D. FOX, P. HAGSTROM, M. MCGINNIS, and D. PESETSKY (eds.) (1998) *Is the Best Good Enough? Papers from the Workshop on Optimality in Syntax*. Cambridge, MA: MIT Press.

BEHRENS, Leila, and Dietmar Zaefferer (eds.) 2002. *The Lexicon in Focus. Competinion and Convergence in Current Lexicology*. Frankfurt am Main: Peter Lang.

CRUSE, D. A. 1986. *Lexical Semantics.* Cambridge: Cambridge University Press.

DO, Anna H.-J., LAURA DOMÍNGUEZ & AIMEE JOHANSEN (eds.) 2001. *BUCLD 25: Proceedings of the 25th annual Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press.

GRIMSHAW, Jane & VIERI SAMEK-LODOVICI 1998. Optimal subjects and subject universals. In: Barbosa et al. (eds.) 1998: 193-219.

LEGENDRE, Géraldine 2001. An Introduction to Optimality Theory in Syntax. In: Géraldine Legendre, Jane Grimshaw & Sten Vikner (eds.) *Optimality-Theoretic Syntax*. Cambridge, MA: MIT Press.

NIEMEIER, Susanne, and RENÉ DIRVEN (eds.) 2000. *Evidence for linguistic relativity*. Amsterdam/ Philadelphia: John Benjamins.

PAPAFRAGOU, Anna, CHRISTINE MASSEY & LILA GLEITMAN 2001. Motion Events in Language and Cognition. In: Do, Anna H.-J. et al. (eds.) 2001: 566-574.

PUSTEJOVSKY, James 1995. *The Generative Lexicon.* Cambridge, MA: MIT Press.

_____. 2002. Polysemy and Underspecification. In: Behrens and Zaefferer (eds.) 2002: 187-208.

QUIRK, Randolph, SIDNEY GREENBAUM, GEOFFREY LEECH & JAN SVARTVIK 1985. *A Comprehensive Grammar of the English Language*. Harcourt: Longman.

RAPPAPORT HOVAV, Malka, and BETH LEVIN 1995. The Elasticity of Verb Meaning. In: Proceedings of the Tenth Annual Conference of the Israel Association for Theoretical Linguistics and of the Workshop on the Syntax-Semantics Interface: 153-171.

SHOPEN, Timothy (ed.) 1995 *Language typology and syntactic description*. Cambridge: Cambridge University Press.

SLOBIN, Dan I. 2000. Verbalized events: A dynamic approach to linguistic relativity and determinism. In: S. Niemeier and R. Dirven (eds.) 2000: 107-138.

TALMY, Leonard 1985. Lexicalization patterns: Semantic structure in lexical forms. In: T. Shopen (ed.) vol. 3: 57-149.