

Semantics of Portions and Partitive Nouns for NLP

Salvador Climent

Facultat de Filologia, Secció Lingüística General
Universitat de Barcelona
Gran Via 585
08007 Barcelona
Spain
climent@goliat.upc.es

Abstract

This paper describes a system of representation of nouns denoting portions, segments and relative quantities of entities, in order to account for this case of part-whole relationship. The semantics of both constructions denoting portions and nouns used to build them are discussed and eventually formalised in a unification-based formalism (LKB-I.RL) in terms of Pustejovsky's Theory of Qualia and Jackendoff's Conceptual Semantics.

1. Introduction and Motivation

Processes of inference of meaning concerning part-whole relations can be drawn from a lexicon bearing meronymic links between words (cf. [WAL88]) e.g. component-whole links (*hand-arm*), member-collection (*tree-forest*) and so. The case of portions, segments and relative quantities of objects or substances (*slices, lumps, buckets, spoonfuls, grains, tops* or *tips* of things) is the exception, since this is a relation which is encoded in the language by means of constructions. Contrary to that which some authors posit ([CIA88], [WIN87]), it doesn't seem to be a productive linguistic generalisation to set in a lexicon some part-of link between *slice* and *cake*. In any case, such relation would exist between *cake* and *slice of cake* -namely, the part-of relation should stand between *slice* and any sliceable thing.

Generally speaking, the relation *portion* may be characterised, semantically, in that the whole pre-exists the part and the part retains most of the properties of the whole ([IRI88]), and, syntactically, as surfacing in $\langle X \text{ of NP} \rangle$ structures, being potentially $\langle X \rangle$ an apparently heterogeneous collection of words which, henceforth, following [COP92] We will refer to as Partitive Nouns (PNs).

On the other hand, portions denoted by such constructions differ from their wholes in some aspects, basically individuation, quantity, process of bringing about, and shape. Such properties, since they are present in $\langle X \text{ of NP} \rangle$ but they were not in $\langle NP \rangle$, are assumed to be carried to the construction by the noun ($\langle X \rangle$).

I'll try to show here that it is plausible to give a global account of such heterogeneous set of words, since they bear a range of common and distinctive linguistic features, and I'll try to provide a representation feasible for NLP which account both

for PNs as a general class and for the homogeneous subclasses which within them could be distinguished and defined¹.

2. Framework

We assume some general familiarity with the framework We work within: LKB-I.RL ([COP92]) as a formalism of computational representation, The Generative Lexicon ([PUS95]) as a theoretical ground for the formalisation of lexical signs, and Conceptual Semantics ([JAC91]) for the conceptualisation of parts and boundaries of entities.

LKB-I.RL is an HPSG-like formalism based on typed feature structures ([CAR92]) developed within the Acquilex Project which makes use of unification and mechanisms of default inheritance. There are four basic levels of representation: ORTH, CAT, SEM and RQS. CAT encodes syntax using a categorial grammar approach, so there are simple (e.g. noun: N) and complex (e.g. adjective: N/N) categories. It is to be noticed that in complex categories the active element is not a category but a lexical sign, in a way that selective restrictions may be set at this level by specifying the semantic type or other features of the selected sign. SEM encodes logical semantics. RQS is basically equivalent to Pustejovsky's Qualia Structure (henceforth QUALIA).

The Theory of Qualia is embedded in that of the Generative Lexicon and has as a basic aim to improve compositionality by endowing lexical signs with a rich internal structure which accounts for different facets of human knowledge about lexical-semantic content, in a way that, in many cases, it allows for avoiding listing separate entries to account for lexical polysemy. [PUS95] posits the lexical entry to be constituted by four structures: Event, Argument, Lexical-Inheritance and QUALIA. The latter consists of four Quales: Agentive (origin or 'bringing about' of objects), Telic (purpose and function), Formal (that which distinguishes it within a larger domain: magnitude, shape, dimensionality) and Constitutive (relation between an object and its constituents or

¹ This work has been developed for Spanish. Notwithstanding, for ease of exposition, I'll exemplify the discussion by means of examples of English -when possible. In any case, for what concerns us here, Spanish and English are, both semantically and structurally, strongly comparable -the kernel $\langle \text{PN of NP} \rangle$ comes out in Spanish as a literal translation, $\langle \text{PN de NP} \rangle$.

proper parts).

[JAC91] posits that speakers understand and schematise a range of phenomena such as mass-count distinction, plurality, partitive constructions and boundary words in terms of elementary conceptual features such as dimensionality, idealised boundaries and internal structure. He introduces features [\pm B] (boundaries being or not in view or of concern) and [\pm I] (entailment or not about internal structure) for both objects and events as expressing the generality of phenomena of boundedness and plurality -the notion of \pm I is present in one or another way in [TAL78] (cf. [JAC91]), [KRI87] (cf. [COP92]) and [LAK87]; that of boundedness in [LAN91]. This feature system comes out as:

+B -I: individuals (*a pig*)/closed events (*John ran to the store*)

+B+I: groups (*a committee*)/bounded iterative events (*the light flashed until down*)

-B-I: substances (*water*)/ unbounded homogenous processes (*John slept*)

-B+I: aggregates (*buses, cattle*)/ unbounded iterative processes (*the light flashed continually*)

Moreover, he posits that boundaries of entities are conceptualised as minimal elaborations (notated *e*) of their ideal topological boundaries (i.e. *the tip of the tongue* is the portion which corresponds to a minimal elaboration of the 0-dimensional point which would ideally be the tongue's tip).

3. Portions and Partitive Nouns

A portion designates a bounded region within a domain ([LAN91]), hence the portion is an individuated entity (even in the case the whole be a substance or mass). The syntactic effect is that, as pointed out in [VOS94], the construction which denotes the portion is syntactically countable.

A portion, an individuated (bounded), object has a shape different from that of the whole. This information is contributed to the construction by the PN. There are PNs which clearly specify shape (RODAJA, *lump*) while others underspecify it (*fragment*). In many cases, PNs, as adjectives do, predicate properties of the portion, specially shape (the translation of Spanish RODAJA must be the paraphrase *round slice*; a *lath* of anything is saliently elongated) or size, but also thickness, consistency or others (as in the case of MENDRUGO, equivalent to *portion* (of bread) except for the fact that entails that the bread is not fresh but stale).

A portion always conveys a measure with relation to the total magnitude of the whole. Therefore, nouns such as *bucket, slice, lump* or *grain* are relative quantifiers in the sense of [LAN91]:

a relative quantifier is so-called because it specifies a quantity in relation to a reference mass; in the default-case interpretation, this reference mass consists of the maximal instantiation of the pertinent category (i.e. its full extension in all conceivable worlds).

A portion has been obtained by a different process than the whole: a cake has been obtained by baking it,

but a slice of cake by cutting it off the cake. The meaning of portions is many times tightly related to such agentive process -if one has obtained 'slices' it necessarily has been by cutting something; there even exists the verb *to slice*. As pointed out in [PUS95], knowledge of 'coming into being' is a property of nominals which has remarkable linguistic consequences since it can account for a reduction of lexical polysemy. For instance, the Spanish phrase HACER PAN [*to-make bread*] means baking bread, whilst HACER REBANADAS (DE PAN) [*to-make slices (of bread)*] means slicing (bread). This way, the very same verb HACER shows two radically different meanings, which in principle should be listed separately in the lexicon. Nevertheless, both can be accounted for in a single entry which selects the Agentive Role of the complement.

More specifically talking about the lexical signs we call PNs, [LYO77] shows that they correspond to the classifiers in languages as Tzeltal, Mandarin Chinese or Vietnamese. In languages with classifiers, these words, semantically strongly similar to determiners and quantifiers, have functions of individuation and enumeration, making surface notions such as sort of entity, shape or measure. E.g. the Chinese phrase SAN BEN SHU is translatable by *three plane-entity book*; *three whiskies* would be constructed with a mensural classifier, being the translation paraphrasable by *three unit/doses whisky*. [LYO77] makes notice that words such as *head (of cattle)*, *sheet (of paper)* or *lump (of sugar)* stand for exactly the same function as classifiers in those languages. Specifically, *lump* operates simultaneously as both a mensural (meaning conventional dose) and classal (denoting a certain type of aggregate) classifier.

Some have assumed that PNs select mass nouns (*slice of cake, glass of wine*), being mass nouns the way in which substances typically surface in the language. Instead, we posit that PNs select both kinds of nouns (count or mass) denoting both kinds of things (individuals or substances), but in any case, crucially, surfacing as expressing cumulative reference.

Let's consider RODAJA DE LIMON [*round-slice of lemon*]. To assume that LIMON here is mass entails assuming that it has undergone a derivational 'grinding' rule which converts countable individuals in masses. Nevertheless, a round-slice of lemon is always a slice of some individual lemon, not a special measure of substance which some time in the past was 'lemon'. In any case, if a 'lemon' weren't an individuated and bounded thing, it couldn't be sliced and the shape of the portion wouldn't depend on that of the whole. The confusing point is that LIMON in the example, RODAJA DE LIMON, surfaces grammatically as substances usually do -namely, zero-determined. But zero-determination is not exclusively a resource to refer to substances, it is the way of expressing cumulative reference. Both individuals and substances may be referred to cumulatively, that is, be construed as an indifferentiated amassment. This surfaces in the language as a zero-determiner plus the

noun in singular in the case of substances (*a glass of wine*), and either in singular or plural in the case of individuals (*a slice of lemon, a basket of lemons*).

So, in our point of view PNs typically select the noun of the whole as it surfaces when construing cumulative reference -but this doesn't compulsorily entails neither the referent is a substance nor it is referred to by means of a mass noun. In the case of individuals, referents still are bounded things, hence both they can be sliced and the shape of their portions still can depend on that of the original whole. We can't go further with the issue here but, at last, what the discussion above stands for is that human conceptualisation is considered as the cause, and the mass-count distinction, as the surface effect.

PNs are not straightforwardly referential, as they predicate with reference to another entity. This may be noticed considering sentences such as *??John ate a slice* or *??I draught three cups*. They are semantically uncomplete as they don't allow the hearer to retrieve from them the information the speaker wants to convey. Further information as in *John ate a slice of cake* or *It was an excellent coffee. I draught three cups* is needed to do the task. When appearing in the discourse, PNs need of further specification of the referent, either via *of*-complementation or via ellipsis or anaphora. Consequently, they can not be unary but relational predicates in the sense of [LAN91], that is, terms which are predicates only with reference to some other entity.

A basic linguistic feature of PNs is that they, as relational predicates, bear selectional restrictions. Namely, each kind of PN can combine with certain referential nouns but can not combine with others, depending on certain features of the referential noun. These features are mostly linguistic (type, countability, singular or plural) but also can depend on knowledge of the world (physical state, etc.). We hypothesise that, in general, distinctions between classes of PNs concerning selectional restrictions must be due to linguistic reasons, while further specifications within each class would be due to properties of the referent. E.g., it could be assumed that containers (*cups, baskets*) select [-B] items (substances and plurals), and more specifically, *cups* select liquids and *baskets* non-liquids.

4. Sorts of PNs

We are not committed here to represent individuation of groups (team, committee) or aggregates (cattle, furniture). In our background ontology, which is intended to represent things in the world as conceptualised by humans (fig.1), these constitute a different class since, in this class of words, speakers' conceptualisation focuses not on the element entities but in their aggregation. Plurals are considered as representing an aggregation (therefore a -B+I concept) of bounded things (therefore either individuals or groups) by means of a derivative (lexical) rule applied on signs denoting those elements (fig.2) - e.g. cow --> cows, team --> teams; but gold,cattle --> *golds/*cattles).

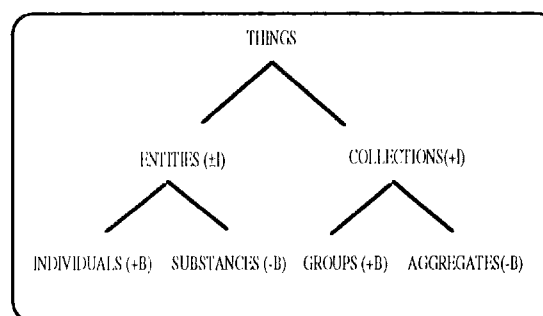


Figure 1. Ontology

Entity-portioning terms have been sorted as follows:

Contents (henceforth CONT) (*a bucket of water, a basket of wheat, a basket of lemons*). They are metonymies of containers expressing a conventionalised measure or quantity of a [-B] entity. Shape is not relevant -a bathtub may contain a bucket of water without there being any bucket in it ([LAN91])

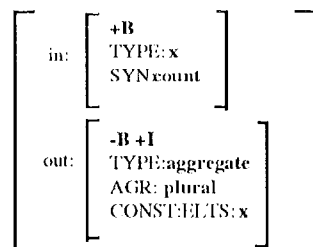


Figure 2. Pluralization (Lexical) Rule

Elements (ELT) (*UN GAJO DE LIMON [an 'inner-pre-existent-division-of-some-fruits'of lemon], a grain of rice*). They are individuations of pre-existing parts of the whole. They select [+I] entities, either individuals or substances. They are not committed to an Agentive process as they may remain attached to the whole. Shape and measure, if considered relevant, are inherent to the portion itself.

Boundaries (BOUND) (*the tip of the tongue, the surface of the sea, the top of a box*). They are idealisations of physical boundaries of the whole. Select [+B] entities. They are non-Agentive either. Their shape is tightly related to that of the whole but one of their dimensions is conceptualised as close to non-existence ([JAC91]). Analogously, they denote a minimal quantity of the whole.

Both **Detached** (DTCI) (*a slice of cake, a slice of lemon*) and **Modelled** (MDLD) portions (*a lump of sugar, a sheet of paper*) have been drawn out of the whole and bear a shape straightforwardly determined by such Agentive process. The meaning of the former focus on the Agentive, which is a process of cut or fragmentation of a [+B] solid. The latter focus on shape, which is often conceptualised schematically (*sheet: a plane; lump, ingot: brick-shaped*). The Agentive is a process of modelling of a [-B] substance.

Summarising about selectional restrictions of these sorts of PNs, they come out as follows:

ELT: select [+I] entities (=individuated or substance) (*a 'pre-existing-inner-division' of LIMON, wheat, sugar*)

BOUND: [+B] entities (=individuated) (*a 'boundary'of*

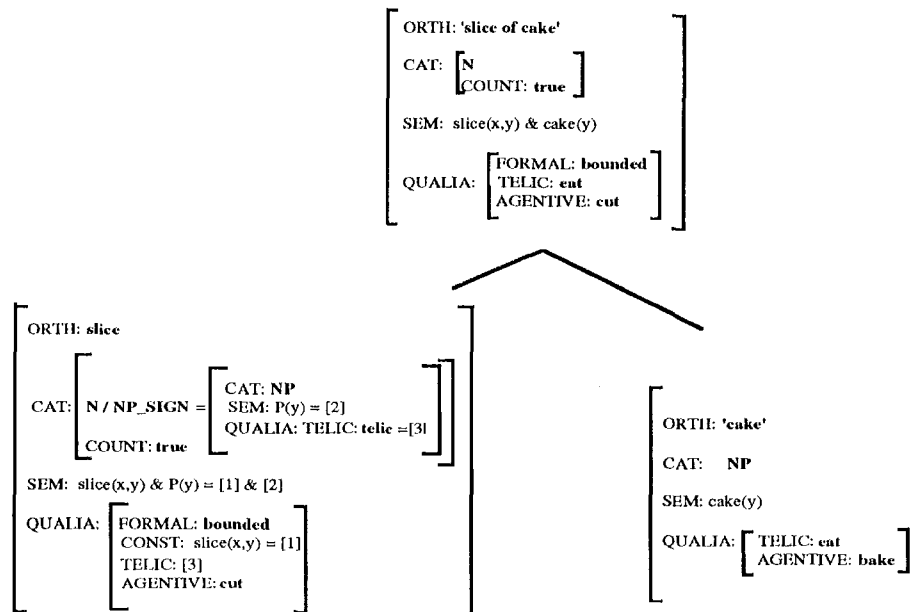


Figure 3. Composition

a tongue, box, cake, LIMON)

CONT: [-B] things (=substances, plurals/aggregates) (*a 'container' of wheat, sugar, water, paper, cakes, tongues, boxes*)

DTCHD: [+B] entities (=individuated) (*a 'detached portion' of a lemon, cake*)

MDLD: [-B] entities (=substances) (*a 'shaped mass' of sugar, wheat, paper*)

With respect to shape, it has to be noticed that while that of **ELT** and **MDLD** is inherent to the portion itself (in **ELT** because the portion pre-existed as an individual; in **MDLD** because the whole was an amorphous mass and it is the process of portioning what has bounded the new thing), in **BOUND** and **DTCHD** shape is somehow relative to the whole. This way, a **RODAJA** is round because it is a cross-cut of either approximately-spherical (*lemon*) or cylindrical (*sausage*) objects; a *slice of bread* will be elliptic or square depending on whether the 'bread' is the classical loaf or the modern polyhedric-shaped one; *top of a box* will show identical behaviour.

Something similar happens with relative quantification. While the measure conveyed by **CONT**, **ELT** and **MDLD** is absolute, that of **BOUND** and **DTCHD** is relative: a *top of a box* or a *slice of bread* will be bigger or smaller depending on the magnitude of the box or the loaf of bread.

5. Representation

To represent PNs in the LKB we have made some interpretations for **FORMAL** and **CONST** Quaales of the **QUALIA**.

We assume that the minimal and most basic **FORMAL** distinction among entities (as conceptualised by speakers) is that of their boundedness or not in terms of [JAC91]. Therefore, this Quale will be assigned to one of both types (or a coherent subtype). Similarly, the minimal constitutive distinction to be done is assumed to be that of entailment or not about internal structure of

things. Therefore, the **CONST** role will be assigned to one of both [\pm I] types (**i_str_true**/**i_str_false**). In this way, selectional restrictions of PNs will be basically defined as selection of signs bearing appropriate types for their **FORMAL** and **CONST** Quaales.

As defined in [PUS95], **SHAPE** and magnitude (**MAGN**) are features of the **FORMAL** Role. Their values can be, as discussed above, either absolute or relative, depending on the kind of portion. In the latter cases, **SHAPE** and **MAGN** of portions will be a function of the corresponding values of the **QUALIA** of the whole². This interpretation of the **MAGN** feature accounts for the nature of relative quantifiers of PNs.

At their turn, [+I] **CONST**s bear the feature **ELTS** e.g. **rice:QUALIA:CONST:ELTS:grain**. (vid. [JAC91], also [COP92]).

[PUS95] (assuming [VIK94] interpretation) attributes the **CONST** Role the function of representing the predicative relation **Part-Of**. Coherently, we assume **CONST** as encoding a relational predicate $R(x,y)$, being R a type taxonomically daughter of **Part-Of**. In the default case, R will be 'Portion' and in more fine-grained cases, a daughter type of it -e.g. 'Slice'.

The Logical Semantics of PNs (**SEM**) will account for their both partitive and relational nature by adopting as predicate-argument structure that of their **CONST** Role. For the sake of unification, in the LKB, **SEM** will be the conjunction of this predicate and the **SEM** value of the sign denoting the whole. This way (vid. fig.3), that of 'slice of cake' will result in **SLICE(x,y) & CAKE(y)** -thus accounting for the logics of *John eats a slice of cake* as **EAT(e,John,x) & SLICE(x,y) & CAKE(y)** versus that of *John eats cake* as **EAT(e,John,x) & CAKE(x)**.

² [JAC91] develops a method to formalise relative shapes including judgements about dimensionality.

For ease of exposition, the syntax (CAT) of PNs is represented here as the complex category N/NP_sign, where the NP_sign is appropriately

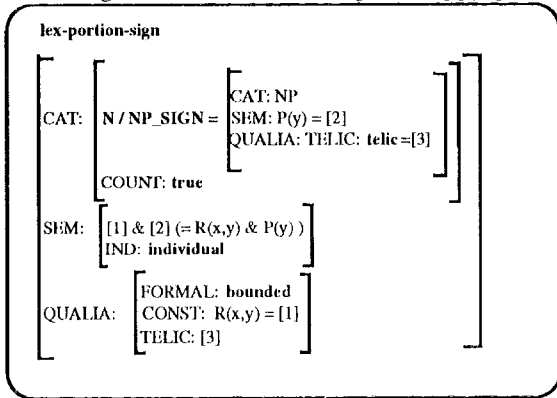


Figure 4. General Portion Sign

specified to account for selectional restrictions and transitivity of properties between the whole and the portion via feature reentrancies. This way, the composition of *slice* (N/NP) and *cake* (NP) will result in an N (*slice of cake*). This accounts for the fact that partitive constructions (e.g. *slice of cake*) do bear the combinatorial possibilities of nouns, while those of PNs are distinct and specific. The preposition (*of*) is omitted here -not in the LKB implementation- since it is irrelevant as it lacks semantic content. We assume so [CHO81] analysis which considers *of* in this kind of constructions a mere surface case-marker. This view is confirmed by data of Spanish (UN VASO VINO = UN VASO DE VINO, UN TROZO PAN = UN TROZO DE PAN) ['a glass wine' = 'a glass of wine', 'a portion bread' = 'a portion of bread'] in which the preposition is elidable -something which is not possible in the case of contentful prepositions (VENGO DE BARCELONA-> *VENGO BARCELONA) ['I-come from Barcelona' -> *'I-come Barcelona'].

Moreover, CAT bears the feature COUNT standing, as well for ease of exposition, for the range of (surface) grammatical behaviour of lexical signs usually referred to as countability/uncountability (see discussion above).

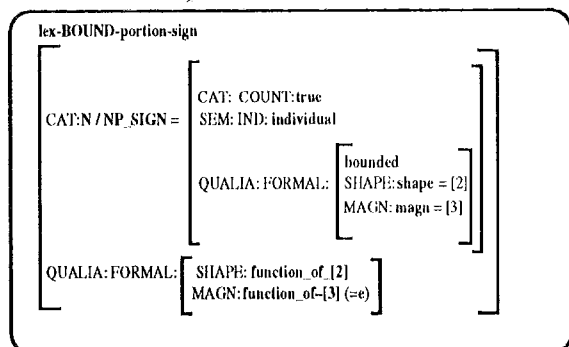


Figure 5. Boundary Portions

Provided all which has been discussed up to here, the general **lex-PORTION-sign** is defined as in **fig.4**; that is, as selecting NPs and resulting in FORMAL:[+B] entity-denoting signs (therefore individuated and syntactically countable) where the only QUALIA feature which percolate from the whole

is the TELIC Role -the rest of Qualess may be overridden by that of the PN.

Specific sub-types are shown in figs. 5 to 9. BOUND PNs (**fig.5**) select [+B] entities (therefore individuated and countable) (*the tip of the tongue*). Being [+B], such wholes bear definite shape and magnitude; therefore such values for the portion will be functions of those of the whole. MAGN, specifically, is assumed to be conceptualised as somehow minimal, therefore notated 'e' ([JAC91]).

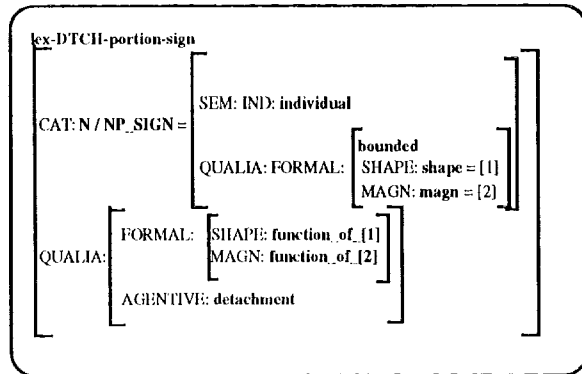


Figure 6. Detached Portions

Similarly, SHAPE and MAGN of DTCHID PNs (**fig.6**) are functions of SHAPE and MAGN of the wholes they select. The difference with BOUND PNs is that the MAGN value doesn't tend to be minimal. The AGENTIVE Role is here relevant -as it as well is in MDLID PNs.

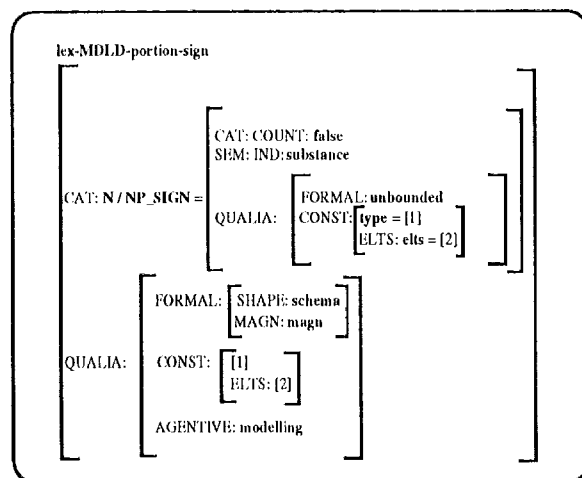


Figure 7. Modelled Portions

What is more remarkable in MDLID PNs (**fig.7**) it that the CONST type of the whole (therefore its value for CONST:ELTS) is inherited by the portion -e.g., if 'sugar' is [+1] and consists of grains, a *lump of sugar* so; if 'paper' has no entailment about internal structure, a *sheet of paper* has not either. MAGN is absolute (i.e., not related to that of the whole, since that is [-B]; e.g., in broad outline, 'all the sugar/paper of the world'). SHAPE is assumed to be schematic (vid. §4).

SHAPE and MAGN of ELT portions (**fig.8**) are also non-relative. These PNs select [+I] entities, usually substances ('wheat', 'rice') but also possibly individuals ('lemon', as conventionalised in Spanish as internally-structured in GAJOS). The value for CONST:ELTS of the whole will be the CONST

predicate of the portion, thus its SEM predicate -e.g.
rice:QUALIA:CONST:ELTS:grain=>
grain:QUALIA:CONST:grain(x,y) =>
grain-of-rice:SEM:grain(x,y) & rice(y).

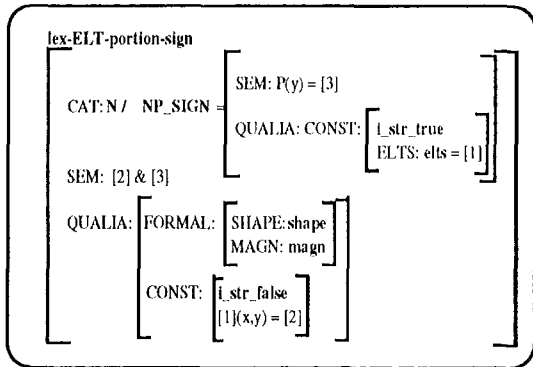


Figure 8. Element Portions

Last, CONT PN's (fig.9) select [-B] items (therefore substances but also plurals)³. The portion retains the constitution of the whole. As discussed above, SHAPE is not relevant, measure (MAGN) is. AGENTIVE, if considered relevant, will be a process of filling the container.

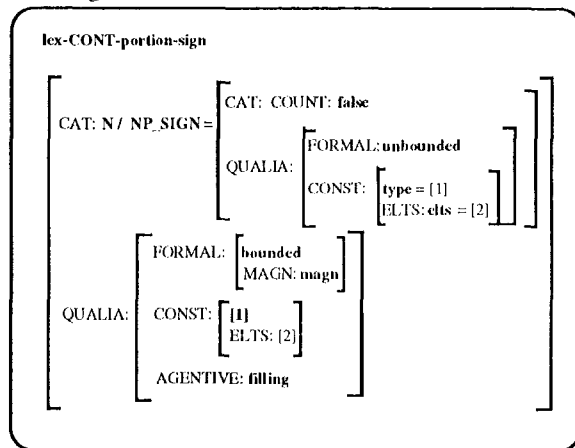


Figure 9. Container Portions

6. Concluding Remarks

In this paper we presented a system of representation of relational nouns denoting portions, segments and relative quantities of entities according to the Theory of Qualia ([PUS95]), which has been partially re-interpreted and adapted to the insights of Conceptual Semantics ([JAC91]). This system accounts for a range of linguistic facts, being the most remarkable the following:

1. Portions are mainly encoded in the language by means of constructions instead of by single lexical units
2. Portions are both bounded entities and relative measures of the wholes of reference
3. Portions inherit from their wholes their purpose or function, but, on the contrary, they show distinctive shape, magnitude and origin

³ Possibly also [-B] collections (groups) (*a wagon of cattle*), but, as said before, we're not committed here to discuss individuation of collections.

4. Partitive Nouns (PNs) select whole-denoting nouns according to the properties of the referent wholes as conceptualised by speakers

PNs have been classified according to the semantic regularities they show, resulting in a system of five basic types, which have been represented in a unification formalism, LKB-LRL ([COP92]), allowing for composition of PN's (e.g. *slice*) and NP's (e.g. *cake*) (plus the semantically void preposition 'of'; in Spanish, DE) into portion-denoting signs (e.g. *slice of cake*).

Acknowledgements. I'd like to thank Drs. Toni Martí and Horacio Rodriguez for their comments and support. This research is funded by the project PB-94-0830 of the DGICYT.

References

- [CAR92] CARPENTER B. 1992. The Logic of Typed Feature Structures. Cambridge University Press.
- [CHA88] CHIAFFIN R. and HERRMAN D.J. 1988. The nature of semantic relations. In Relational Models of the Lexicon, Martha Walton Evens ed. Cambridge U. Press
- [CHO81] CHOMSKY N. 1981. Lectures on Government and Binding. Foris, Dordrecht.
- [COP92] COPESTAKE, A. 1992. The Representation of Lexical Semantic Information. Ph. D. Thesis. University of Sussex (Cognitive Science research paper CSRP 280)
- [IRI88] IRIS M.A., LITOWITZ B.E. and WALTON EVENS M. 1988. Problems of the part-whole relation. In Relational Models of the Lexicon, Martha Walton Evens ed. Cambridge University Press.
- [JAC91] JACKENDOFF R. 1991. Parts and Boundaries. In Lexical and Conceptual Semantics, Levin & Pinker eds. Elsevier Science Publishers. Amsterdam.
- [KRI87] KRIFKA, M. 1987. Nominal Reference and Temporal Constitution: Towards a Semantics of Quantity. Proceedings of the 6th. Amsterdam Colloquium pp. 153-173. University of Amsterdam.
- [LAN91] LANGACKER R. 1991. Foundations of Cognitive Grammar. Vol II. Descriptive Application. Stanford University Press. Stanford, CA.
- [LAK87] LAKOFF G. 1987. Women, Fire and Dangerous Things: What Categories Reveal about the Mind. University of Chicago Press. Chicago.
- [LYO77] LYONS J. 1977. Semantics. Cambridge University Press. Cambridge.
- [PUS95] PUSTEJOVSKY J. 1995. The generative Lexicon. The MIT Press. Cambridge, Ma.
- [TAL78] TALMY, L. 1978. The Relation of Grammar to Cognition: A Synopsis. In D. Waltz (Ed.), Theoretical Issues in Natural Language Processing 2. New York: Association for Computing Machinery.
- [VIK94] VIKNER C. and P. HANSEN. 1994. Knowledge Bases and Generative Lexicons. Proc. of 'Workshop on Lexical Semantics', University of Copenhagen.
- [VOS94] VOSSEN P. & COPESTAKE A. 1994. Untangling Definition Structure into Knowledge Representation. In Briscoe, de Paiva, Copestake Eds. "Inheritance, Defaults, and the Lexicon", CUP, Cambridge.
- [WAL88] WALTON EVENS M. Ed. 1988. Relational Models of the Lexicon. Cambridge University Press.
- [WIN87] WINSTON M., CHIAFFIN R. and HERRMAN D. 1987. A Taxonomy of Part-Whole Relations. Cognitive Science 11, 417-444.