UNIVERSIDADE DE LISBOA

FACULDADE DE LETRAS

DEPARTAMENTO DE LINGUÍSTICA GERAL E ROMÂNICA

COMPUTATION OF VERBAL PREDICATES IN PORTUGUESE: RELATIONAL NETWORK, LEXICAL-CONCEPTUAL STRUCTURE AND CONTEXT

THE CASE OF VERBS OF MOVEMENT

Raquel Amaro

DOUTORAMENTO EM LINGUÍSTICA

LINGUÍSTICA COMPUTACIONAL

2009

UNIVERSIDADE DE LISBOA
FACULDADE DE LETRAS
DEPARTAMENTO DE LINGUÍSTICA GERAL E ROMÂNICA

COMPUTATION OF VERBAL PREDICATES IN PORTUGUESE: RELATIONAL NETWORK, LEXICAL-CONCEPTUAL STRUCTURE AND CONTEXT

THE CASE OF VERBS OF MOVEMENT

Raquel Amaro

Orientadora
Professora Doutora Palmira Marrafa

Co-orientadora
Professora Doutora Christiane Fellbaum

DOUTORAMENTO EM LINGUÍSTICA
LINGUÍSTICA COMPUTACIONAL

2009

This work was supported by Fundação para a Ciência e a Tecnologia
grant SFRH/BD/13875/2003

a Sérgio Carlos da Fonseca,

músico, pintor e poeta

avô, amigo e herói

Acknowledgements

The first words of this dissertation go to those who made this work possible. To all whose support I sincerely and overtly want to acknowledge, although the simple gesture of registering their contribution cannot express all my gratitude for the help, encouragement and inspiration they provided me.

To Professor Palmira Marrafa, for always teaching so much, for allowing me to benefit from her expertise as a linguist and for her example of commitment to rigorous scientific research; for the always sharp and wise comments and ideas that value this work; for the availability for discussing even the minor details. For the trust, wisdom and true respect with which she works alongside her students. For all the support, generosity and friendship.

To Professor Christiane Fellbaum, who in spite of being so far away, always took the time to comment and discuss the ideas depicted in this work, kindly sharing her insights on the data and the analyses. For her knowledge, encouragement and dedication. For introducing me to Lexical Semantics, and, with her enthusiasm, unknowingly defining such a significant part of my path.

To Sara Mendes, with whom I work side by side in so many projects, for her unreserved support. For always being available to discuss and comment this work, so many times helping me solving its shortcomings, and for so generously volunteering to do its final reading. Above all, for her friendship, solidarity and complicity.

To Professor Fernanda Bacelar do Nascimento, who welcomed me into CLUL, giving me the privilege of working alongside her in a truly exceptional environment. For her knowledge and concern and for all the learning opportunities. And most of all, for the friendship and interest with which she always encouraged me to go further.

To Rui Chaves, for the encouragement and support that helped me taking the first steps in this work. For the friendship and fellowship. For the contagious enthusiasm and devotion to scientific research.

To Amália Mendes, for the way she so naturally inspires others to pursue new and defying challenges, for the enthusiastic way how she shares her knowledge, and for the unwavering friendship and trust.

To Rita Veloso, for so closely sharing this path with me, discussing so many questions and hypotheses. For the unconditional support, friendship and care, and for always sheltering me in my short visits to Lisbon.

To my other colleagues at CLG. To Catarina Ribeiro and Ricardo Santos, for so friendly welcoming me into the group. To Susana Lourosa for the friendship, encouragement and long nights of deep conversations.

To my friends and co-workers at CLUL, Florbela Barreto, Sandra Antunes, Luísa Alice Pereira and Catarina Magro, for the concern, support and good-tempered discussions.

To all those at CLUL who make this institution so much more than just a host institution! For the warmth, support and concern. To João Saramago, Gabriela Vitorino and Luísa Segura, for the kind and always good-humored disposition.

A special thanks to Professor Ana Maria Martins, for taking part in my application to the PhD grant.

To the linguists with whom I've crossed paths and who had the generosity of discussing with me parts of what became this work, for the spontaneous words of encouragement.

To my large, loud and incredible family! To my mother Zaida, for trusting my abilities and continuously encouraging my engagement in new challenges; for all the patience and care. To my grandmother Lucília: por todas as saudades, avó, e por todo o carinho e apoio. To my sisters Rita, Clara and Mariana, for our special complicity: for the care, faith and availability to help no matter what. To Ana, André, Eddie, Fernando, Rui, Sara and Víctor, for always taking interest, for all the support and for the noisy and fun reunions and discussions. To D. Adelaide, Sr. Gil, D. Maria, Sr. Joaquim, Lina, Zé Miguel, Mariana and Laurentino, for truly welcoming me into the family, so many years ago. And, naturally, to my nephews, António, Nuno, Ricardo, Joana and little João, for all the laughs, playdates and contagious genuine joy.

To my amazing friends Catarina, Fernando, Paula, Marco, Graça, Ricardo, Paula and Mariana for blindly believing in the conclusion of this work since day one. For the care and encouragement, and for all the stress-free and fun Friday night evenings.

And to Nuno. For always being there: in the ups and downs and in-betweens. For embracing my path as his own. For the unconditional support, care and love which I cherish everyday. For all the joyfulness, patience and trust, and for peacefully enduring my many absences. And, obviously, for the music: lots and lots of music!

Thank you all for granting me this extraordinary learning and living experience!

Resumo

Inserida no campo da Semântica Lexical Computacional, e com base no pressuposto de que o desempenho de processos computacionais de determinação do significado beneficia grandemente do uso de recursos lexicais extensos e estruturados, esta dissertação apresenta uma análise de verbos de movimento do Português, com o objectivo de determinar as propriedades semânticas e sintácticas destes itens lexicais e a forma como esta informação se relaciona com a computação e previsão das estruturas em que estes verbos podem ocorrer.

A restrição do objecto de estudo a um domínio semântico específico permitiu uma determinação mais precisa do significado de cada verbo, através do estabelecimento de relações léxico-conceptuais num modelo relacional do Léxico. A análise da semântica lexical destes verbos tem como base as especificidades de significado que diferenciam os verbos hipónimos dos seus hiperónimos e dos seus nós irmãos. A identificação de componentes do significado partilhados e não partilhados por verbos de um mesmo domínio semântico motiva a definição da informação semântica relevante a representar ao nível da entrada lexical, bem como a determinação da estrutura desta informação.

No âmbito deste trabalho, é apresentada uma proposta de wordnet de verbos de movimento, referindo os diferentes níveis de análise relevantes para uma representação coerente dos verbos desta classe: a forma como os itens lexicais são agrupados em conjuntos de sinónimos que denotam conceitos e as relações estabelecidas entre estes conjuntos contemplam as propriedades conceptuais e semânticas dos itens lexicais, e a organização do léxico daí resultante permite determinar qual informação partilhada.

A construção de uma wordnet de verbos de movimento do Português impôs a definição dos nós de topo da rede, bem como a determinação de outras opções de codificação, permitindo testar a herança conceptual pelos nós mais baixos da hierarquia. A rede obtida revelou a diversidade semântica e sintáctica de verbos directamente relacionados, e, particularmente, que propriedades semânticas, tais como a estrutura argumental ou propriedades de Aktionsart, estão directamente relacionadas com a especificação dos conceitos denotados, mas não são directamente herdadas ou condicionadas pelo domínio semântico a que um dado verbo pertence.

Com base na wordnet desenvolvida, é apresentada uma análise decomposicional do significado dos verbos de movimento do Português, evidenciando as especificidades de significado que diferenciam os nós hipónimos dos seus hiperónimos. Esta análise revelou padrões de incorporação semântica diferentes dos descritos por Talmy (1985) para as línguas românicas, e resultou na proposta de um novo conjunto de componentes semânticos, lexicalizados nos verbos estudados, mas extensível à análise de verbos de outros domínios semânticos.

O conteúdo semântico específico de cada verbo hipónimo diferencia verbos co-hipónimos e explica a incompatibilidade entre co-hipónimos: são incompatíveis (i.e., não co-ocorrem) co-hipónimos que lexicalizam valores opostos, ou de outro modo incompatíveis,de um mesmo componente semântico.

A lexicalização dos componentes semânticos considerados afecta em vários graus a herança de propriedades do hiperónimo, nomeadamente no que respeita a propriedades relativas à estrutura argumental (número de argumentos, propriedades de subcategorização e restrições semânticas do tipo de argumentos seleccionados) e a propriedades de Aktionsart.

Foram observados os seguintes padrões de lexicalização: a incorporação de restrições relativas aos componentes semânticos ORIGEM (local ou posição inicial) e DESTINO (local ou posição final) resulta no aumento do número de argumentos seleccionados sintacticamente realizados, ao passo que a lexicalização destes componentes resulta na diminuição do número de argumentos sintacticamente realizados, comparativamente com a estrutura argumental do hiperónimo. A lexicalização de TRAJECTO (localizações intermédias entre a ORIGEM e o DESTINO) resulta no acréscimo de mais um argumento, relativamente à estrutura argumental do verbo hiperónimo, tipicamente correspondendo a um argumento que denota OBJECTO DE REFERÊNCIA (objecto externo relativamente ao qual o evento é perspectivado), realizado sintacticamente na posição de objecto; a incorporação de restrições a este componente semântico (TRAJECTO) resulta no aumento do número de argumentos seleccionados sintacticamente realizados e reflecte-se na selecção de um argumento sintacticamente realizado, denotador de TRAJECTO do evento de movimento, introduzido pela preposição *por*.

As alterações de propriedades de Aktionsart na wordnet de verbos de movimento do Português, i.e., hipónimos com valores de Aktionsart diferentes dos dos seus hiperónimos, ocorrem com a lexicalização de DESTINO e ORIGEM. A lexicalização destes componentes resulta em eventos de tipo accomplishment ou achievement, dado que a definição da localização ou posição final (DESTINO) ou da localização ou posição inicial (ORIGEM) estabelece um limite ao evento, transformando um evento de tipo actividade num evento de tipo accomplishment ou achievement.

A representação dos itens lexicais aqui proposta é feita no quadro do Léxico Generativo (LG) e contempla três níveis de representação distintos: a estrutura argumental, a estrutura eventiva e

a estrutura qualia. Os itens lexicais estão, por sua vez, integrados numa estrutura de herança lexical.

De forma a conseguir uma caracterização mais completa dos verbos de movimento do Português, especificamente no que diz respeito às suas propriedades de subcategorização, é proposta a modelização de preposições na WordNet.PT (WN.PT) e a sua representação lexical no quadro do LG. A integração das preposições na WN.PT segue investigação existente sobre modelos ontológicos de representação de preposições, nomeadamente no que toca aos conceitos denotados por estes itens lexicais, consensualmente adoptados quer pelas gramáticas tradicionais, quer análises linguísticas actuais. Esta integração resulta num tratamento coerente e uniforme de preposições semanticamente plenas, que introduzem argumentos verbais, mas também de preposições marcadoras de argumento.

Através da utilização dos níveis e elementos de representação do LG, é proposta a representação integral de verbos de movimento do Português, dando conta da percolação de informação no léxico, do impacto da lexicalização de componentes semânticos nas propriedades semânticas e sintácticas dos verbos e da compatibilidade entre co-hipónimos.

A utilização recursiva das estruturas lexicais disponíveis permite a percolação da informação através das redes de hiperonímia e possibilita uma codificação coerente e económica da informação, incluindo propriedades de subcategorização significativas. As estruturas lexicais resultantes mostram como a relação de hiponímia pode substituir redes ortogonais de tipos, no que respeita ao estabelecimento e à definição das propriedades semânticas através de estratégias de subtipificação. Além disso, a permeabilidade ao contexto de que dão conta os mecanismos generativos integrados no LG, em particular os mecanismos de subespecificação e de co-composição, assegura a plasticidade que explica a diversidade de comportamentos sintácticos dos itens lexicais, directamente relacionada com as suas propriedades léxico-semânticas.

Para a definição de um léxico computacional que modelize as propriedades semânticas e sintácticas dos itens lexicais é proposta a integração das estruturas informacionais do LG nas wordnets: as estruturas informacionais do LG permitem entradas lexicais estruturadas e o modelo da WordNet, pela sua natureza, fornece a necessária hierarquia lexical que permite o acesso a outras estruturas no léxico. A integração dos níveis de representação do LG, nomeadamente da estrutura argumental, da estrutura qualia e da estrutura eventiva, prova que as wordnets podem comportar descrições lexicaisde maior granularidade, que suportam o tratamento de vários fenómenos léxico-conceptuais, sem comprometer a sua arquitectura.

A integração de informação relativa à estrutura argumental na WN.PT é conseguida através da implementação de três novas relações: a relação SELECCIONA/É SELECCIONADO POR; a relação INCORPORA/É INCORPORADO POR e a relação SELECCIONA POR DEFEITO/É SELECCIONADO POR DEFEITO POR. A integração da estrutura qualia é obtida pela associação de relações léxico-conceptuais aos

papéis qualia, sem qualquer perda de informação, no que constitui um processo simples e económico. A expressão da estrutura eventiva no modelo da WordNet, por sua vez, é alcançada através de um novo conjunto de traços (Tipo de evento, Argumentos, Subeventos, Restrições e Núcleo) que permite a associação das propriedades internas dos eventos aos synsets e a sua codificação na base de dados. A representação sistemática de informação relativa à estrutura eventiva, para além de permitir a descrição da ordem dos argumentos, enriquece o poder descritivo destes recursos.

A integração dos níveis de representação do LG em wordnets tem como resultado repositórios de informação semântica lexical mais ricos e estruturados que contemplam informação relativa aos papéis qualia e que permitem a extracção de informação relativa às estruturas argumentais e eventiva dos itens lexicais, ou seja, léxicos generativos sobre os quais podem operar mecanismos como a co-composição, a ligação selectiva e a coerção de tipos.

As propriedades semânticas e sintácticas consideradas nas entradas lexicais dos verbos analisados fornecem também pistas para dar conta de restrições de ocorrência destes verbos em algumas construções. Dando particular atenção à selecção de argumentos denotadores de local e de OBJECTO DE REFERÊNCIA, realizados sintacticamente na posição de objecto, à expressão de movimento direccionado em Português, à ocorrência de verbos de movimento em construções médias e não-causativas e à distribuição do clítico *–se* nestas construções, este trabalho apresenta também a análise dos diferentes comportamentos linguísticos dos verbos de movimento do Português nestes contextos e a relação destes comportamentos com as propriedades léxico-semânticas dos verbos.

Apesar de não permitir um tratamento exaustivo de todos os comportamentos observados, a caracterização léxico-semântica proposta neste trabalho constitui um passo necessário para permitir o tratamento dos fenómenos observados, avançando algumas explicações que permitem dar conta destes diferentes comportamentos.

Verbos que lexicalizam ORIGEM e DESTINO ou TRAJECTO seleccionam objectos que denotam OBJECTO DE REFERÊNCIA, i.e., argumentos verdadeiros que denotam entidades concretas e delimitadas, expressos sintacticamente por SNs.

A possibilidade de ocorrer em estruturas de movimento direccionado, i.e., com SPs que expressam a ORIGEM e o DESTINO do movimento, está directamente relacionada com as propriedades semânticas e sintácticas dos verbos analisados: verbos de mudança de localização legitimam e/ou restringem a ocorrência destes constituintes, de acordo com os componentes semânticos lexicalizados e com as suas propriedades de subcategorização. Ainda no que diz respeito à expressão de movimento direccionado em Português, os dados analisados mostram que a distribuição dos verbos de movimento do Português com SPs denotadores de DESTINO introduzidos pela preposição *a* é condicionada pelo tipo de evento de movimento denotado pelo verbo (modo de movimento vs. movimento direccionado), mas também pelas propriedades de

Aktionsart dos verbos, uma vez que os SPs introduzidos por *a* induzem uma interpretação pontual do estado final do evento, refutando assim as análises de verbos de movimento nas línguas românicas baseadas apenas nas restricções de ocorrência destes verbos com esta preposição.

A correlação entre a proeminência de uma causa externa ou agente e a impossibilidade da sua ocorrência em construções não causativas dá conta da distribuição dos verbos de movimento nestas construções: verbos que lexicalizam INTENÇÃO ou um componente de MODO forte que implique a acção de uma causa externa ou agente não entram em construções não causativas.

A análise da distribuição do clítico *–se* em construções médias, não causativas e passivas levou levantou a hipótese de o clítico induzir uma interpretação de envolvimento de um actor externo no evento: as construções passivas pressupõem necessariamente uma causa externa, logo exigem o clítico; nas construções médias o clítico marca os casos em que há a pressuposição do envolvimento de um actor externo no evento; e nas construções não-causativas, o clítico marca a correlação entre o agente e o tema/paciente do evento, forçando uma leitura não-causativa com sujeitos sintácticos [-animados].

Neste trabalho, fica patente que a modelização dos itens lexicais de uma dada categoria gramatical não é independente da de itens de outras categorias com que estes podem ocorrer, o que, necessariamente, aumenta o escopo da nossa análise. Para além disso, fica demonstrado que a modelização dos itens lexicais no modelo da WordNet compreende uma estrutura de herança lexical motivada, permitindo uma descrição adequada e económica dos itens lexicais e potenciando a construção de recursos lexicais de grande escala para fins computacionais.

Abstract

Within the field of Computational Lexical Semantics, and based on the assumption that the performance of meaning determination computational processes is largely assisted by structured and extensive lexica, providing different types of information, this dissertation presents the analysis of Portuguese verbs of movement in order to determine the semantic and syntactic properties of these lexical items and how this information can be related to the computation and prediction of the structures in which they occur.

The restriction to a specific semantic domain allowed a more accurate determination of the meaning of each verb, through the establishment of lexical-conceptual relations within a relational model of the Lexicon. The lexical semantic analysis of these verbs is based on the meaning specificities that differentiate hyponym verbs from their hyperonyms and sister nodes. The identification of the meaning components shared and those not shared by verbs of the same semantic domain motivates the determination of the relevant semantic information to be stated at the lexical entry level, as well as the structure of this information.

This work puts forth a proposal for a Portuguese wordnet of verbs of movement, referring the different levels of analysis that are relevant for a coherent encoding of the verbs of this class: the way lexical items are grouped in concept denoting sets and the relations established between these sets contemplate the conceptual and semantic properties of the lexical items, and the resulting organization of the lexicon allows for the determining the information that is shared.

The development of a wordnet for Portuguese verbs of movement required the definition of the top nodes of the net as well as of some other coding options, allowing testing conceptual inheritance from the higher to the lower nodes in the hierarchy. The resulting network revealed the semantic and syntactic diversity of verbs directly related, namely that semantic properties such as argument structure or Aktionsart properties are directly related to the meaning specificities of the concepts denoted, but are not straightforwardly inherited or conditioned by the semantic domain to which a given verb belongs.

Based on the developed wordnet, a decompositional analysis of the meaning of the Portuguese verbs of movement is presented, focusing on the meaning specificities that differentiate each hyponym concept with regard to its hyperonym. This analysis revealed semantic incorporation patterns different from those considered to work for Romance languages and resulted in the proposal of a new set of semantic components, comprising the elements lexicalized by the verbs in study, and extendable to the analysis of verbs from other semantic domains.

The semantic content specific to each hyponym differentiates co-hyponym verbs and explains co-hyponyms compatibility: co-hyponyms lexicalizing opposite or otherwise incompatible values for the same semantic element are incompatible (i.e., do not co-occur).

The lexicalization of the semantic components considered affects the inheritance of the hyperonym properties at different degrees, namely in what concerns argument structure (argument number, subcategorization properties and semantic restrictions on the type of the arguments selected) and Aktionsart properties.

The following salient patterns of lexicalization were observed: the incorporation of restrictions on the semantic components SOURCE (initial location or position) and GOAL (final location or position) results in an increase of the number of overt arguments of the hyponyms, whereas the lexicalization of these components results in a decrease of the number of overt arguments of the hyponyms, with respect to the hyperonym argument structure. The lexicalization of PATH (medium locations between the SOURCE and the GOAL) results in the increase of one more overt argument to the argument structure of the hyperonym verb, usually corresponding to a GROUND (external object with respect to which the event is put in perspective) argument realized in object position; the incorporation of restrictions on this semantic component results in the increase of the number of overt arguments, reflected in the selection of an overt argument referring the PATH of the movement event and is introduced by the preposition *por* (through).

Aktionsart shifts within the wordnet of Portuguese verbs of movement, i.e., hyponyms that display Aktionsart values different from those of their hyperonyms, occur with the lexicalization of GOAL and SOURCE. The lexicalization of the elements SOURCE and GOAL result in accomplishment or achievement type events, since the determination of a specific final location or position (GOAL) or initial location or position (SOURCE) establishes a limit to the event, shifting an activity type event to an accomplishment or achievement type event.

The lexical items representation is done within Generative Lexicon (GL) framework and contemplates three distinct levels – argument structure, event structure and qualia structure. Lexical items are integrated in a lexical inheritance structure.

In order to better characterize the Portuguese verbs of movement, specifically in what concerns subcategorization properties, the modelization of prepositions in WordNet.PT (WN.PT) and their semantic representation at the lexical entry level in the GL framework, is proposed. The integration of prepositions in WN.PT follows previous research on ontological models for the

representation of prepositions, namely in what concerns the concepts denoted by prepositions consensually adopted in traditional grammars and state of the art models. This results in a coherent and unified treatment of the semantically full prepositions that introduce verbal arguments but also of argument-marking prepositions.

Using these levels and elements of representation, a complete representation of Portuguese verbs of movement is proposed, accounting for the percolation of information within the lexicon, for the impact of semantic lexicalization in the semantic and syntactic properties of verbs and for verbal co-hyponym compatibility.

The recursive use of available lexical structures allows the percolation of information through the hyponymy trees and enables a coherent and economic codification of the information, including significant subcategorization properties. The resulting lexical structures demonstrate that hyponymy can replace a semantic type lattice in what concerns establishing and defining semantic properties by subtyping strategies. In addition, the permeability granted by the GL model principles, in particular underspecification and co-composition, assures the necessary context flexibility to explain the diversity of syntactic behaviors directly related to lexical semantics properties.

For the definition of a computational lexicon that models the semantic and syntactic properties of lexical items, the integration of informational structures in wordnets is proposed: GL lexical structures provide the structured lexical entries, and WordNet, by its nature, provides the necessary lexical hierarchy that conveys the access to other structures in the lexicon. The integration of GL representation levels in a wordnet, namely argument structure, qualia structure and event structure, demonstrates how wordnets can support a finer-grained lexical description that provides the bases for accounting for several lexical semantic phenomena, without compromising the architecture of the model.

The integration of argument structure information in WN.PT is achieved through the establishment of three new relations: SELECTS/ IS SELECTED BY relation; INCORPORATES/IS INCORPORATED IN relation and SELECTS BY DEFAULT/IS SELECTED BY DEFAULT BY relation. The integration of qualia role in wordnets is attained by associating lexical-conceptual relations to qualia roles, without any loss of information, in what consists of a simple and low cost process. The expression of event structure in wordnets is accomplished through a new set of features (Event type, Arguments, Subevents, Restrictions and Head) that encode the internal properties of the events. The systematic representation of event structure information, besides providing the grounds for argument order description, enriches the descriptive power of these resources.

The integration of GL representation in wordnets results in richer and more structured repositories of lexical semantic information that contemplate qualia information and allow the extraction of argument structure and event structure information, i.e., generative lexica over which devices such as co-composition, selective binding and coercion can operate.

The semantic and syntactic properties considered in the lexical entries of the Portuguese verbs of movement also provided insights on the occurrence restrictions displayed by these verbs in some constructions. Focusing on the selection of arguments denoting location and GROUND occurring in object position, the expression of directed motion in Portuguese, the occurrence of verbs of movement in middle and non-causative constructions and the distribution of *–SE* in these constructions, this work also presents the analysis of the different behaviors of Portuguese verbs of movement in these contexts and their relation with the lexical semantic properties of the verbs.

Although not accounting exhaustively for all the different behaviors observed, the lexical semantic characterization proposed constitutes a necessary step to enable the treatment of the observed phenomena and provides some explanations of different behaviors.

Verbs that lexicalize SOURCE & GOAL or PATH select defined GROUND objects, i.e., true arguments denoting concrete and bounded entities, syntactically expressed by NPs.

The possibility of occurring in directed motion structures, i.e. with PPs that express the SOURCE and GOAL of the movement is directly related to the semantic and syntactic properties of the verb at stake: verbs of change of location license and/or restrict their co-occurrence with these constituents, according to the semantic elements lexicalized by the verbs and to their subcategorization properties. Regarding also the expression of directed motion in Portuguese, the data show that the distribution of Portuguese verbs of movement with GOAL denoting PPs introduced by the preposition *a* (roughly corresponding to the English preposition *to* in some contexts) is conditioned by the type of movement event denoted by the verb (manner of motion vs. directed motion), but also by Aktionsart properties, since PPs introduced by *a* induce a punctual aspect interpretation of the final state of the event, and refute the analyses of verbs of movement in Romance languages based solely on the co-occurrence restrictions with the preposition *a*.

The correlation between the prominence of an external cause or agent and the impossibility of occurring in non-causative constructions accounts for the distribution of verbs of movement in these constructions: verbs that lexicalize INTENTION or a strong MANNER component implying the action of an external cause or agent do not enter non-causative constructions.

The analysis of the distribution of *–se* in middle, non-causative and passive constructions lead to the hypothesis of the *-se* inducing the interpretation of the involvement of an external actor in the denoted event: passives with *-se* necessarily entail an external cause and thus require the presence of the clitic; in middle constructions, the clitic marks the case where the involvement of an external actor in the denoted event is entailed; and, in non-causative constructions, the clitic marks the correlation between the agent and theme/patient participants of the event, forcing the non-causative reading with [-animated] syntactic subjects.

From this work, it is apparent that the modeling of lexical items of a given POS is not independent from that of others of different POS with which they may occur, which necessarily extended the scope of the analysis depicted here. Moreover, it is demonstrated that modeling lexical items in the WordNet model, establishing a motivated lexical-conceptual inheritance structure, allows for an an economic and adequate description of lexical items and potentiates the construction of large-scale lexical resources suitable for computational purposes.

Table of contents

1. Introduction

During the course of the seminars for my master's degree I was introduced to Lexical Semantics and got immediately captivated by it. Lexical Semantics studies the meaning of words – defined as intuitive units of language, roughly corresponding to sequences of characters comprised between spaces in written languages, or, in some cases, groups of these sequences that convey a given concept – and how this meaning can be represented.

It was this particular point that caught my attention: if, in a first and naïf perspective, it seems somewhat trivial to determine the meaning of a word (we use words daily to communicate therefore we must have a solid idea of what they mean), it is not easy to accurately and formally represent this information. This task gets even more complex within any generative approach where, desirably, the information gathered to represent the meaning of words should also account for their meaning in context, i.e. serve the purpose of computing the meaning of whole sentences.

The ascertainment of the core meaning of a given word, the direct interaction between lexical semantics and syntax (given the fact that it is almost impossible to isolate the study of the semantic content of a given word from its syntactic realization) along with the element of formal representation are, in my perspective, the main appealing features of this field of Linguistics.

My continuous involvement in projects within the scope of Computational Lexical Semantics contributed to the determination of the general and base idea for the current dissertation: accounting for meaning in context requires large lexica, with functional content, covering much more than one specific set of words defined for a given study.

The computational processes of meaning determination are largely assisted by extensive lexica, organized, and with different types of information. The lexicon is seen as a complex knowledge system, crucial to the processing of language, and not a static repository of irregular information. The organization of the lexical items within the lexicon and the type of information that is stated in the lexical entries play, therefore, an important role in the computation of meaning and cannot be dissociated from semantic and syntactic processing issues.

Also, researchers on Natural Language Processing (NLP) are, nowadays, well aware of the need for semantic specification in computational lexica and are confronted with issues such as sense identification and disambiguation, the degree of granularity of semantic information, length and organization of the lexicon, and so on. Aspects such as these have been in the path of many linguists, concerned with the representation of semantic information in the Lexicon, required for natural language processing, an area less addressed by earlier research, which focused mainly on syntactic issues. Consequently, the Lexicon acquired a relevant position within linguistic theories and models and Lexical Semantics has seen its role enhanced.

According to the literature[1], the information in the Lexicon must consider the polymorphic properties of language and the creative use of words, namely phenomena such as polysemy, compositionality and context sensibility, thus requiring a coherent treatment and analysis of lexical semantics. This implies that the information in the lexical entries should provide insights on how and why the interaction of the meaning of words works and when it is productive, and also leads to the need of taking into account the syntactic structures in which lexical items occur.

The demand for useful computational lexica, however, is far greater than the current offer, since the construction of such resources requires fundamental linguistic research to determine the informational content of the lexical entries, as well as formal modeling of the syntactic and semantic properties of the lexical items for meaning computation purposes, also ideally mirroring the organization and properties of the mental lexicon.

The work presented here is thus strongly motivated by its potential contribution to the fulfillment of this gap. Our research is dedicated to the analysis of the lexical-conceptual structure and organization of verbs of movement in Portuguese, the semantic and syntactic properties of these verbs, as well as the emerging linking patterns from lexical semantic properties to syntax, aiming at the construction of a verbal lexicon that, besides modeling linguistic knowledge, and thus reflecting concerns regarding complexity, also considers the representation of information useful for computational purposes.

The remainder of this chapter presents the object of this dissertation: the verbs of movement in Portuguese. Section 1.1 sums up the goals of this work. Section 1.2 provides some insights on the choice of the verbal category and of the semantic class of verbs of movement in particular, as well as the main issues relevant for their treatment. The directions taken in this research, directly related to the framework adopted as well as to the goals pursued, are presented in section 1.3, and, finally, section 1.4 presents the outline of this dissertation.

[1] See Pustejovsky (1991, 1993, 1995), Fellbaum (1998a, 1998b, 1999), Buitelaar (1998), Fong *et al.* (2000), Dang *et al.* (2000), for instance.

1.1 Goals

The construction of a lexicon for computational purposes desirably reflects the principles that govern the structure and organization of the concepts denoted by the lexical items, as well as the lexical-syntactic mapping patterns that may arise. The choice of the model of the lexicon, the determination of the pertinent information in the lexical entries, as well as its modeling and structure, and the organization of lexical items within the lexicon are necessarily related to this task.

Assuming the construction of a lexicon for meaning computation purposes and the fundamental research required for achieving it as our goals, we focus on the verbal lexicon, specifically on the class of verbs of movement. The choice of this particular set of verbs is related to two major issues, described in the next section: verbs of movement are a quite consensually defined class that exhibits semantic and syntactic diversity.

On the one hand, restricting our object to a specific semantic domain allows a consistent representation of the items that compose it in a relational model of the lexicon, in which lexical items are linked by lexical-conceptual relations. Verbs of movement are hierarchically related and, thus, the meaning components shared and not shared among them can be more easily identified.

On the other hand, this unity is not enough to explain the diversity of syntactic behaviors that might be predictable in a deeper lexical semantics analysis. The determination of the relevant semantic information and its structure may prove to be crucial steps for the explanation and prediction of the syntactic structures in which these verbs occur.

Verbs of movement have been the object of several studies, in particular focusing on English data. However, the conclusions of the aforementioned studies do not accurately account for the Portuguese data. As it will be shown further ahead, there are several constructions described for English in which Portuguese verbs of movement do not occur, whereas there are others whose acceptability differs in the two languages. This way, the study depicted here has also a contrastive character by addressing phenomena that are common in both languages and phenomena that are unique to Portuguese, and possibly other Romance languages, such as the clitic SE distribution in non-causative and middle constructions.

The different syntactic constructions in which verbs of movement occur are common to other verb classes in English as well as in Portuguese. Thus, the determination of the restrictions that condition these constructions is expected to extend to other classes of verbs, resulting in a potential universal treatment of several phenomena.

In sum, our purpose is to construct a relational lexicon for verbs of movement that can be the base for an accurate computation of meaning, including meaning in context, providing semantic grounds for the syntactic behavior of these lexical items. Moreover, the relational design of the lexicon assumes a hierarchical structure that combined with an inheritance device allows for an adequate description of lexical items and may constitute a solid starting start for the construction of a lexicon for computational purposes.

Assuming a mixed approach combining a conceptual and a syntactic analysis of lexical units, and profiting from two distinct frameworks, we also aim at providing a valid contribution to the fundamental linguistic research in progress within the field of Computational Lexical Semantics, considering that the traditions of either focusing on syntactic behavior, on the one hand, or on the modeling of meaning through world models, on the other, are not necessarily competing.

> "The central role of the syntax-driven lexical semantics in the process of deriving the meaning of a text is to decode the nature of the dependency between heads of phrases and their arguments in a particular language. This knowledge is then used in ontology-driven lexical semantics as a necessary set of heuristics which allow us to represent the meaning of a text in terms of language-independent conceptual models. Thus, we believe that a comprehensive approach should combine the benefits of both approaches and that neither will on its own be sufficient for realistic NLP."
>
> Nirenburg & Levin (1991:6)

1.2 Verbal lexicon

As predicates, verbs establish relations between entities, events and situations, being their argument selection universally considered an intrinsic part of their meaning. This implies that more or less subtle meaning differences of a given verb can be related with, and/or reflected in, the number and the semantic and syntactic natures of its arguments (see (1)). Also, the semantic properties of a given verb necessarily condition the adjuncts with which it can co-occur, as exemplified in (2) below.

(1) a. The bartender drinks **soda**. (*drink* ≅ ingest liquids)

b. **The bartender** doesn't drink ∅. (*drink* ≅ drink alcoholic beverages)

c. **Fish** don't drink. (*drink* ≅ ingest liquids)

(2) a. The man ran. (*run* ≅ ACTIVITY-type event)

b. The man ran **for hours**. (*run* ≅ ACTIVITY-type event)

c. The man ran **into the house**. (*run* ≅ ACCOMPLISHMENT-type event)

d. The man ran **3 miles**. (*run* ≅ ACCOMPLISHMENT-type event)

This context "sensibility" poses several difficulties on the treatment of the verbal lexicon but constitutes clear evidence that the semantic description of the items of this lexical category cannot be dissociated from their syntactic realization.

For this reason, verbs constitute a good starting point for the construction of the lexicon:

i) for their intrinsic characteristic of predicates, verbs enter a large diversity of semantic and syntactic phenomena, requiring thus complex informational structures;

ii) the semantic and syntactic information stated in a verbal lexical entry is necessarily connected to and/or conditioned by the information stated in the lexicon for other POS.

Verbs can thus illustrate, to a great extent, the structure and informational content of the lexicon, given that the establishment of the relevant semantic and syntactic information for a given verb lexical entry also depends on the available informational structure and content of its arguments, therefore contributing to the determination of the relevant information for other POS entries and to the overall design of the lexicon.

1.2.1 Verb classes

The compartmentalization of the lexicon into classes – according to syntactic parameters or to semantic and conceptual properties – is a well-known approach that intends to echo generalizations concerning the syntactic behavior or the semantic properties of sets of lexical items. Classes represent domains within which lexical items share syntactic and/or semantic features and constraints, such as argument type and adicity, diathesis alternations, and so on.

The verbal lexicon can be clustered into classes according to several types of properties: syntactic properties (transitive verbs, intransitive verbs, inchoative verbs, etc.), Aktionsart properties (state verbs, process or activity verbs, accomplishment or telic process verbs, semelfactive or atelic punctual situation verbs and achievement or punctual situation verbs) and semantic or conceptual properties (change of state verbs, psychological verbs, sound emission verbs, verbs of movement, and so on). These approaches are not exclusive, some properties being reflected and conditioned by others.

Although it is clear that the syntactic description of a given lexical item is of the utmost importance for the processing of language structures (it allows to discard impossible syntactic constructions besides providing a base or prototypical syntactic construction for a given item, among other things), the classification of verbs based only on syntactic features such as subcategorization structure or syntactic realization establishes a limited number of classes, which do not reflect the crucial similarities and differences among the lexical items clustered, as exemplified in (3), (4) and (5) below.

(3) Transitive verbs (verbs that require an object as complement): *eat*, *build*, *frighten*, *suffer*, …

(4) Indirect transitive verbs (verbs that subcategorize a PP as object): *suffer* (from), *think* (about), *depend* (on), *object* (to)…

(5) Unaccusative verbs (verbs that realize the theme argument in the subject position): *come*, *bloom*, *disappear*, …

The classification of verbs according to the type of event they denote, regarding the internal and aspectual properties of these events, results also in a very small set of classes: state verbs (*be*, *know*, *sound*,…), activity verbs (*run*, *read*, *eat*, …), accomplishment verbs (*build*, *bake*, *climb*), semelfactive verbs (*cough, knock, sneeze*, …) and achievement verbs (*recognize*, *find*, *start*,…)[2]. Although reflecting several objective characteristics of the members of each set, these classes do not reflect all the syntactic and semantic properties that differentiate or associate the verbs of a given language. Also, the base Aktionsart properties of a given verb can be altered in context, by morphological tense and aspect marking or by the interaction with other phrases (see Moens (1987)), without corresponding to different lexical items:

(6) a. *read* (≅ ACTIVITY-type event, ongoing)

b. The man **has read** the bible. (≅ completed event, ACCOMPLISHMENT)

c. The man read the bible **in twelve hours**. (≅ completed event, ACCOMPLISHMENT)

(7) a. *build* (≅ ACCOMPLISHMENT-type event, completed)

b. The man **is building** the house. (≅ ongoing event, ACTIVITY)

As stated before, the determination of classes of lexical items based on syntactic, conceptual or semantic properties are not incompatible. Much research work[3] strongly supports the relation and integration of the various aspects of linguistic knowledge, namely syntactic and semantic properties, in lexical modeling, instead of providing an isolated and independent treatment of each separate aspect.

[2] Examples taken from Smith (2004).

[3] Atkins *et al.* (1986), Pustejovsky (1991, 1993, 1995), Kilgarriff (1993), Levin (1993), Marrafa (1993), Saint-Dizier (1995), Buitelaar (1998), Fellbaum (1998a, 1998b, 1999, 2003), Kipple & Gurney (1998), Fernández *et al.* (1999), Lapata & Brew (1999), Sag & Wasow (1999), Agirre & Martinez (2000), Dowty (2000), Fong *et al.* (2000), Krifka (2001), Mendes (2001), Piñon (2001), Fong & Fellbaum (2003), among others.

For this reason, the classes of verbs currently proposed usually reflect the different types of information associated to lexical items, specially the ones concerning syntactic and conceptual properties.

1.2.1.1 Syntax-driven semantic classes

Levin (1993) is a noteworthy example of how to use the syntactic behavior of verbs as a key grouping factor, based on the hypothesis that the behavior of a given verb is the reflex of its meaning. Verbs are grouped according to the alternations patterns they exhibit. The methodology proposed is based on the assumption that diathesis alternations reflect semantic properties. Thus, verbs that exhibit the same diathesis alternation patterns share some aspect of meaning. Levin (1993) presents the set of diathesis alternations that occur in English and the classes of verbs obtained by the observation of the occurrence of verbs in pairs of alternative constructions. The final result is a study of classes of English verbs that share both meaning and syntactic features.

Although quite exhaustive, this pattern recognition methodology raises three main issues: (i) it does not offer explicit answers on how to identify the relevant meaning components that are responsible for the diathesis alternations and patterns (see Levin 1993:14); (ii) it is based on the diathesis alternations in which English verbs participate, which may not correspond straightforwardly to the alternations available in other languages, rendering impossible to apply the same methodology universally; (iii) lexical-conceptual similarities and/or dissimilarities are not necessarily mirrored in the resulting classes.

The first issue is directly related to the descriptive character of the work and reflects only the acknowledgement of the research yet to be done, which is, definitely, largely facilitated by such an exhaustive groundwork.

The second one results from cross-linguistic comparison and does raise questions regarding the universal character of the verb classes identified, at least in what concerns the grouping factors established. It seems difficult to determine classes of verbs in one language, based on alternations that are not available for the verbs in that language. The examples in (8) show some of the alternations in English that do not occur with the correspondent verbs in Portuguese.

(8) a. Induced action alternation
The horse jumped over the wall./The man jumped the horse over the wall.
O cavalo saltou por cima do muro. /#O homem saltou o cavalo por cima do muro.

b. Conative alternation
The man pushed the table./The man pushed at/against/on the table.
O homem empurrou a mesa./*O homem empurrou à/contra a/na mesa.[4]

c. Dative alternation
The man brought the dress to Ana./The man brought Ana the dress.
O homem trouxe o vestido à Ana./*O homem trouxe a Ana o vestido.

The occurrence of verbs in the conative alternation, for instance, is mentioned in Levin (1993: 136, 137-138) as a factor that allows for distinguishing *Carry* Verbs from Verbs of Exerting Force: verbs from the first class do not allow this alternation and verbs from the second class do. A verb such as *push* is listed as a member of both classes, its sense allowing or disallowing the alternation, or, conversely, the occurrence in the conative alternation conditioning its sense. The straightforward application of this criterion to the correspondent Portuguese verb *empurrar* (push) would result in its inclusion in the *Carry* Verbs class alone, although this verb can also be used to convey the notion of exerting force:

(9) a. O homem empurrou a porta com toda a sua força para que esta não fechasse.
(The man pushed (at) the door with all his strength so that it wouldn't close.)

Also related with this issue is the question of how to consider the verbs occurring in alternative constructions, when these constructions result in meaning changes. Take, for instance, the locative preposition drop alternation, in (10). The absence of the preposition does not result in a merely alternative construction in Portuguese, since the PP is used to convey "Path" or "Means", whereas the NP conveys "Obstacle" (Fong & Fellbaum 2003), thus more co-occurrence restrictions seeming to be involved.

(10) Locative preposition drop alternation
a. O homem desceu $\underline{\text{pelas escadas(*todas)}}_{\text{PATH}}$/ $\underline{\text{pelo elevador(*todo)}}_{\text{MEANS}}$.
(The man descended $\underline{\text{by (all) the stairs}}_{\text{PATH}}$/ $\underline{\text{by (all) the elevator}}_{\text{MEANS}}$)

b. O homem desceu $\underline{\text{as escadas (todas)}}_{\text{OBSTACLE}}$/* $\underline{\text{o elevador (todo)}}_{\text{OBSTACLE}}$
(The man descended $\underline{\text{(all) the stairs}}_{\text{OBSTACLE}}$/$\underline{\text{(all) the elevator)}}_{\text{OBSTACLE}}$

The third issue is directly related to the observation of the resulting classes of verbs: if, on the one hand, the members of a given class share semantic and syntactic properties, these properties are, on the other hand, transversal between classes. Verbs in different classes can

[4] Some Portuguese verbs of movement seem to enter a similar alternation. For instance:

a. O homem remexeu a gaveta./O homem remexeu na gaveta.
(the man rummaged the drawer/the man rummaged in the drawer)

However, this example is more similar to the spray/load alternation, since the absence of the preposition entails that the entire drawer was rummaged.

occur in the same alternation constructions, but, conversely, verbs conceptually related are distributed by several classes. For instance, *Verbs of Sound Emission*, which include *bark, hiss* or *squeak*, is a particular class and *Verbs of Sounds Made by Animals*, including verbs such as *sing, roar* or *hiss*, is a different class altogether, although many verbs are members of the two classes. The presence of these verbs in both classes is justified by their semantic content and syntactic behavior. And yet, nor the resulting set of classes is organized, nor the semantic relations between the verbs in both classes are explicitly stated.

In sum, the determination of classes according to the alternation patterns exhibited by verbs can provide evidence of their semantic content. However, it is still necessary to look deeper in order to determine the semantic elements that may be responsible for different syntactic behaviors, even among verbs of the same class.

> "The key to maintaining this hypothesis [of semantic determination] is the identification of the appropriate representation of verb meaning. Determining the appropriate meaning components is not easy, since a priori it is possible to classify verbs in many ways according to their meaning."
>
> Levin (1993:13)

Note that conceptual grouping factors are also quite visible in syntax-driven semantic classes, even if not directly assumed. Levin's verb classes are to some extent conceptually determined: we have *Verbs of Perception*, *Verbs of Existence*, *Verbs of Communication*, *Verbs of Motion*, for instance, that gather verbs that denote events from the same conceptual domain, although, as stated before, some conceptually related verbs fall under different and unrelated classes.

1.2.1.2 Concept and semantic domain based classes

Concept-based classes typically group lexical items of a same semantic domain. Verbs are clustered according to the concept they denote based, for instance, on ontologies of the existing situations that can be classified as events. Concept and semantic domain based classes can be used to determine relevant information to represent the meaning of verbs, namely what distinguishes a given class from other lexical classes, and what distinguishes a given verb from the other verbs in the same class.

According to Miller & Johnson-Laird (1976), the following semantic domains (and respective classes) can cover almost all the verbal lexicon of a language, with the exception of state denoting verbs: verbs of motion, verbs of perception, verbs of contact, verbs of communication, verbs of competition, verbs of change, verbs of cognition, verbs of consumption, verbs of creation, verbs of emotion, verbs of possession, verbs of bodily care, verbs of bodily functions, verbs of social behavior and interactions.

Following the hypothesis that some part of the syntactic and semantic behavior of lexical items in context can be traced back to lexical semantic properties, conceptual and semantic properties constitute a crucial clustering factor for organizing the lexicon.

1.2.2 Verbs of movement

Movement is a core notion of our perceptual experience and conceptualization of reality. The perception of things in motion is one of our first experiences as infants and it seems to be related to our conceptualization structure[5]. It is a concept widely and universally lexicalized in natural languages, which indicates a strong potential applicability of the results of this work to other languages, at least to those languages reflecting a common spatial reckoning (see Levinson (2003)).

1.2.2.1 Conceptual coherence

Verbs of movement convey, more or less specifically, an event occurring in a given time in which something moves from one place to another, as further explained ahead (see representations in (11) and (12)). Langacker (1987, 1991) presents motion as an event composed of a sequence of noncumulative states, s_1, s_2... s_n, in which a moving entity successively occupies a given location, l_1, l_2 ... l_n. "A motion verb can be regarded as a special sort of perfective process, namely one in which each component state specifies the relation between the mover and its immediate location." (Langacker 1991:155). Talmy (1985, 2000b), on the other hand, defines motion events as situations that contain movement or the maintenance of a stationary location (Talmy 1985: 85). A motion event is decomposed into basic semantic components, *figure, ground, path*, and described with regard to a reference frame: "The Figure is a moving or conceptually movable object whose path or site is at issue. The Ground is a reference frame, or a reference object stationary within a reference frame, with respect to which the Figure's path or site is characterized" (Talmy 2000b:26). Lakoff (1987), Johnson (1987) and Fillmore *et al.* (2000), among others, argue that motion events also comprehend other structural components such as source, goal and direction, that together with the notion of figure (or theme), ground (or landmark) and path (or trajectory) compose an abstract schema of motion.

Regardless of the representations of the motion events proposed, and some of their implications in terms of the internal structure of events, all these authors present motion as the core concept at stake. Verbs of movement can, thus, be defined as the class of verbs whose core concept is an event where there is change of spatial location over time. Here we adopt the designation "Verbs of movement", instead of "verbs of motion" or "motion verbs", since we will

5 See Cristobal (2001), Gennari *et al.* (2001), Teixeira, J. (2001), Levinson (2003), among others.

consider two major subsets as natural members of this class: verbs that denote change of location (corresponding more directly to the class of motion verbs) and verbs that denote change of position.

Change of location verbs, a 'location' being the space occupied or able to be occupied by an entity, denote a movement event where the initial location in which an entity starts the movement event is different from the final location in which an entity is when the movement event ends: the entity moves from location A to location B ($\neq$ A):

(11) Change of location event:
For a given entity E; for the locations $L_1, ..., L_n$; at a given time span $\{t_1, ..., t_n\}$;
E changes location iff:
E in t_1 is in L_1,
E in t_n is in L_n, and
$L_1 \neq L_n$.

Change of position verbs, a 'position' being the spatial configuration of an entity within a location, denote a movement event of some part of an entity (that can amount to the whole entity), with respect to locations contained within a reference location: some part of an entity (or all of it) moves within a frame location L, see (12).

(12) Change of position event:
For a given entity E; for the locations $L'_1, ..., L'_n$; at a given time span $\{t_1, ..., t_n\}$;
E changes position iff:
$L'_1, ..., L'_n \subset L$;
E (or part of E) in t_1 is in L'_1,
E (or part of E) in t_n is in L'_n,
$L'_1 \neq L'_n$.

In general terms, change of location verbs involve translational movement and typically can occur with source, path and goal denoting PPs, whereas change of position verbs do not involve translational movement and, therefore, do not typically occur with source, path and goal denoting phrases:

(13) O João desceu do telhado$_{SOURCE}$ para o pátio$_{GOAL}$ pelas escadas$_{PATH}$.
(John descended from the roof$_{SOURCE}$ to the yard$_{GOAL}$ through the stairs$_{PATH}$)

(14) ?*O João abanou o arbusto do jardim$_{SOURCE}$ para a estrada$_{GOAL}$ pelo relvado$_{PATH}$.
(John shook the bush from the garden$_{SOURCE}$ to the road$_{GOAL}$ through the lawn$_{PATH}$)

Note, however, that sentences like the one in (14) can be greatly improved if we consider the specific kind of locations that a change of position event is restricted to: if the frame location is respected, it is possible to refer to other sub-locations in which the entity is located at some point, although it is not certain whether these might be considered as source or goal locations or rather defining of the direction of the movement.

(15) a. O João abanou o arbusto de um lado para o outro (*pelo ar).
(John shook the bush from one side to the other (through the air))

b. A mulher abanava o leque da cara para o peito (*pelo ar).
(The woman shook the fan from her face to her chest (through the air))

The fact that path denoting phrases cannot co-occur with these verbs can indicate that the co-occurring phrases, although pointing to specific locations, are used to describe the direction of the movement event, as for instance, 'from left to right' or 'up and down'. This explanation is reinforced by the strangeness of sentences in which a change of position verb occurs only with one of the denoting locations phrases:

(16) a. ?O João abanou o arbusto de um lado.
(John shook the bush from one side)

b. ?A mulher abanava o leque para o peito.
(The woman shook the fan to her chest)

Asher & Sablayrolles (1996) further subdivide the class of verbs of movement. They propose four major subclasses: i) c*hange of location verbs*; ii) *change of position verbs*, iii) *inertial change of position verbs*; and iv) *change of posture verbs*. Briefly, the authors define these classes with regard to the notions of ***location*** – portion of space denoted by a lexical item, associated with a functionality; ***position*** – portion of surface defined by the 3-D portion of space occupied by an entity, without any associated functionality; and ***posture*** – specific way in which an entity is within the 3-D portion of space it occupies (Asher & Sablayrolles 1996:170). In reality, this foursome classification reflects a subdivision of the class of change of location verbs considered above, since *change of posture verbs* correspond to the defined class of change of position verbs.

Accordingly, *change of location verbs* (ex. *enter, arrive, land, approach*) imply a *strict internal path* and require an argument introducing a location different from the source location. *Change of location* verbs entail that the moving entity changes location during the process (see (17)).

(17) a. O homem entrou na casa.
(The man entered the house)

Change of position verbs (ex. *move around, circulate, descend/go down*) entail that the moving entity changes position during the process, i.e., the entity does not occupy the same portion of surface during the process, usually occur with an argument introducing a background location and cannot occur with phrases such as *in place*:

(18) a. O homem movimentava-se no quintal.
(The man moved around in the yard)

b. *O homem movimentava-se sem sair do mesmo sítio.[6]
(The man moved around in place)

Inertial change of position verbs (ex. *run, dance, fly*) imply, but do not entail, a change of position of the moving entity. Therefore, these verbs can occur with phrases such as *in place*:

(19) a. O homem corria sem sair do mesmo sítio.
(The man ran in place)

In our perspective, the subdivision of the change of location verbs class presented in Asher & Sablayrolles (1996) does not accurately represent these verbs. The division between *change of location verbs* and *change of position verbs* is based on the distinction of two types of spatial entities – surfaces and spaces – which are related to 2 vs. 3 dimensional objects and to the lexicalizations of these objects. Verbs such as *enter* or *exit* seem to require a 3 dimensional location, but not necessarily, since sentences such as *He entered the yard* are possible. Also, some *change of position verbs* can occur with goal denoting phrases:

(20) a. The firemen descended into the basement.

b. The children went up to the attic.

With respect to the distinction between *change of position* vs. *inertial change of position* classes, some verbs of *inertial change of position* result in awkward sentences when occurring with *in place*-type phrases, demonstrating that there might be more than just an implication of motion:

(21) a. ?#The soldiers crawled in place/without leaving their location.

b. ?#The children skated in place/without leaving their location.

Our point here is that, although reflecting some typical properties and co-occurrence restrictions of the verbs in question, the classes established – *change of posture verbs*, *change of location verbs*, *change of position verbs* and *inertial change of position verbs* – are based on non-conclusive tests and do not present new or relevant information. *Change of posture verbs* correspond to the already established class of change of position verbs. In the case of *change*

[6] We chose here to translate 'in place' by 'sem sair do mesmo sítio' (without leaving the same place), since sentences such as 'O homem movimentava-se no mesmo sítio/local' (The man moved around in the same place) can be interpreted as referring to the location in which the movement event takes place.

of location verbs, it seems to us that the entailment of change of location is a direct reflection of the Aktionsart properties of these verbs, which denote a final state. The verbs included in the third class – the so called *change of position verbs* - correspond to directed motion verbs, i.e., verbs that denote a direction of movement, which, in our opinion, seem to entail a change of location – especially since the distinction between location and position is somewhat subtle. Finally, the *inertial change of position* class seems to correspond to the class of manner of motion verbs, which explains the focus on the manner component of the event in detriment of the motion component. Also, the specification of manner of motion can be responsible for the different acceptability of the sentences in (21), above.

In sum, here we will consider two larger verb classes: change of location verbs and change of position verbs, according to the definitions established in (11) and (12), which together constitute the class of verbs of movement, since their core concept concerns a movement event.

In spite of the conceptual universality and coherence of the verbs that integrate the class of verbs of movement, this class often gathers verbs with diverse syntactic and semantic behavior, requiring thus a deeper lexical semantic analysis, and covering a series of phenomena that also occur with other predicates. It is our perspective that the study of verbs requires the treatment of issues transversal to the entire lexicon, constituting thus a good starting point for the current work.

1.1.2.2 Semantic and syntactic diversity

The organization of lexical items within the lexicon and the type of information that is stated in lexical entries desirably reflect the syntactic realizations of lexical items. Interestingly, Portuguese verbs of movement, although part of the same lexical-conceptual class, show, in some cases, different syntactic behaviors. In this section we will present examples of the syntactic diversity observed in verbs of movement in Portuguese in what concerns argument structure, Aktionsart properties and several constructions, namely directed motion constructions, middle constructions and causative/non-causative alternations.

1.1.2.2.1 Argument structure

An adequate model of the verbal lexicon must necessarily deal with predicate valence, reflected in restrictions on argument selection: verbs that select one argument are expected to form a group, verbs that select two arguments another group, verbs that select three arguments yet another group. Besides the number of arguments selected, it is also expected that verbs in the same class impose the same semantic and syntactic restrictions. Portuguese verbs of movement include verbs conceptually and semantically related but showing different selection restrictions. For instance, the verbs *mover* (move) and *tirar* (take), although lexical-conceptually related (see (22)), have a different number and type of arguments, as shown in (23).

(22) a. *mover* (move) ≅ change location
b. *tirar* (take) ≅ move from a given location

(23) a. Ele moveu o caixote.
(He moved the box.)

b. Ele tirou o caixote <u>da rua</u>.
(He took the box <u>from the street.)</u>

Another example of an issue to be considered in the establishment of argument structure for verbs of movement concerns arguments in object position. Usually these arguments denote measure, i.e., the length of the movement event, but can also denote "obstacles" or "regions" (Fong & Fellbaum 2003). According to Fong & Fellbaum (2003), in English, obstacle arguments

ii) refer to bounded areas (# *The fighter planes crossed the air*);

iii) cannot be measured (**John crossed the river for 200 yards*);

iv) can enter (periphrastic) passive constructions (*The bridge was crossed*);

v) can enter the middle construction (*This river crosses easily*); and

vi) allow -*ing* nominalizations (*The crossing of the river was difficult*),

whereas region arguments

ii) refer to unbounded areas (*Mary ambled the countryside*);

iii) can be measured (*Mary ambled the street for 3 miles*);

iv) cannot enter (periphrastic) passive constructions (**The streets were ambled*);

v) do not enter middle constructions (**The new boardwalk ambles easily*); and

vi) do not allow -*ing* nominalization (**?The ambling of the streets was ostentatious*)[7].

In Portuguese, region denoting arguments can also be expressed by PPs and not only by NPs, avoiding the ambiguity between expressions that can constitute obstacle or region arguments, but illustrating different argument structures for semantically close verbs.

(24) a. O João percorreu as ruas de Paris.
(John walked around/toured the streets of Paris)

[7] Examples taken from Fong & Fellbaum (2003).

b. O João deambulou **pelas** ruas de Paris.
'John ambled/strolled through.the streets of Paris'

In what concerns the distinction between region and path denoting arguments in Portuguese, when realized as PPs introduced by the preposition *por* (rouglhly corresponding to the English preposition *through*), the differences are, in most cases, defined contextually. In the sentence (25)a, the PP *pela rua* can be interpreted as a path or as a region - unbounded –, if further specified by a larger context (see (25)b and c). Regions, in Portuguese, can also be unambiguously expressed by PPs introduced by the preposition *em* (≅ in)(see (25)c).

(25) a. O João andou pela rua$_{\text{[path or region]}}$.
(John walked the street/through the street)

b. O João andou pela rua$_{\text{[path]}}$ até à escola.
(John walked through the street to school)

c. O João andou pela/na rua$_{\text{[region]}}$ durante horas.
(John walked the street for hours)

Also, the distinction between obstacle and measure denoting arguments in Portuguese is not straightforward. The following examples show different types of phrases occurring with the manner of motion verb *andar* (walk).

(26) a. O João andou 100 metros/ uns 100 metros/ a rua (toda).
'DET John walked 100 meters/ INDEF.ART 100 meters/ the (entire) street'

b. O João andou metade de 100 metros/ ?de uns 100 metros/ da rua.
'DET John walked half of 100 meters/of INDEF.ART 100 meters/ of.the street'

c. 100 metros/ ?*Uns 100 metros/ ?* A rua (toda) foram/foi andados/a pelo João.
'100 meters/INDEF.ART 100 meters/the (entire) street were/was walked by John'

d. 100 metros/ ?Uns 100 metros /?esta rua (toda) andam/a-se facilmente.
'100 meters/INDEF.ART 100 meters/this (entire) street walk/s easily'

e. a caminhada de 100 metros/?de uns 100 metros/ da rua (toda)
'the walking of 100 meters/INDEF.ART 100 meters/ of.the (entire) street'

According to these examples and to the properties described for obstacle denoting arguments, it is possible to observe that measure arguments have a somewhat hybrid status (see Rappaport Hovac & Levin (2002:5)):

i) measure arguments do not necessarily refer to bounded (but measured) areas; they do not enter the periphrastic passive; and some nominalizations are not possible;

ii) some measure arguments correspond to bounded areas (*a rua* (the street)); can be partitioned; can enter middle constructions; allow nominalization; and establish the end point of the activity denoted by the verb, changing the aspectual value of the sentence to that of an accomplishment.

Although obstacle and measure denoting arguments share many properties, obstacles typically correspond to overt arguments and measures do not, see (27). Also, measure denoting arguments (that refer to the distance covered by the movement event) can measure the event in terms of spatial distance or time, and obstacle denoting arguments cannot, see (28).

(27) a. O João andou. (John walked)
b. *O João atravessou. (John crossed)

(28) a. O João andou 50 metros/ a rua toda/50 minutos/o dia todo.
(John walked 50 meters/the entire street/50 minutes/the whole day)

b. O João atravessou a ponte/*50 metros/*50 minutos/*o dia todo.
(John crossed the bridge/50 meters/50 minutes/the whole day)

The issues illustrated in this section, on argument number and type as well as on the subcategorization properties of the verbs of this class, have to be addressed in order to achieve a lexical semantic representation of these items that accounts for the syntactic diversity observed in this class of verbs, allowing the computation of meaning in context.

1.2.2.2.2 Aktionsart properties

The Aktionsart properties of verbs are not necessarily mirrored in the organization of the lexicon, verbs denoting different types of events being close to each other. For instance, the verb *regressar* (return) is an accomplishment denoting verb, that can be represented as a subtype of the activity denoting verb *mover-se* (move oneself), in (29). As well known, these Aktionsart differences are reflected in co-occurrence restrictions with certain aspectual adverbials: accomplishment denoting verbs can occur with *in*-adverbials but cannot occur with *for*-adverbials, whereas activity denoting verbs can occur with *for*-adverbials but not with *in*-adverbials (see (30)). The verbs *mover-se* and *regressar* are both verbs of movement in a subtyping lexical-conceptual relation, i.e. they share the same core concept, but do not share the same Aktionsart properties.

(29) a. *mover-se* (move oneself) ≅ change oneself's location
b. *regressar* (return) ≅ move oneself back to the start location

(30) a. Ele moveu-se *em meia hora/durante meia hora.
(He moved in half an hour/for half an hour)

b. Ele regressou em meia hora/*durante meia hora.
(He returned in half an hour/for half an hour)

Once again, it is necessary to consider this diversity and, more importantly, to determine to what extent Aktionsart properties have to be considered in the organization of the verbal lexicon. The previous examples clearly show that the lexicalization of events, and especially the co-relation between lexical items denoting similar concepts, does not follow a strict organization of event types, typically mirrored in ontologies.

1.2.2.2.3 Constructions

Other issues arise when considering several constructions in which verbs of movement occur. Based on Levin (1993), we compared the listed constructions in which English verbs of movement (change of location and change of position denoting verbs) occur with the corresponding constructions allowed for Portuguese verbs of movement. The three constructions that raise issues on the semantic and syntactic information in the lexical entries in which Portuguese verbs of movement occur, and, hence, that require further analysis are the directed motion construction, the middle construction and the causative/non-causative alternation. In this section, we will present some of the issues to be considered within the scope of this work.

1.2.2.2.3.1 Directed motion construction

The possibility of occurring with directional structures, i.e. with PPs that express the source, and goal of the movement, is frequent within the class of verbs of movement, although some restrictions of occurrence with these expressions are yet to be accounted.

Gutiérrez (2001) groups the different approaches to this type of structures in the literature, namely in what concerns the information stated at lexical entry level. Briefly, in one approach (Jackendoff (1983, 1990), Rappaport Hovac & Levin (1998), Van Valin & Lapolla (1997), etc.) it is considered that the occurrence with directional expressions reflects meaning differences that are stated in the lexical entry of verbs, since all pertinent information, including the argument structure, is determined by the conceptual structure of the verb. The lexical entry for *run* proposed in Jackendoff (1990:45) is presented in (31)a:

(31) a. *run* (manner of motion verb (activity))
V
$\_\langle \text{PP}_j \rangle$
$[_{\text{Event}}\ \text{GO}\ [_{\text{Thing}}\ ]_i,\ [_{\text{Path}}\ ]_j]$

b. *enter* (directed motion verb (accomplishment))
V
_<NP$_j$>
$[_{Event}\ GO\ ([_{Thing}]_i,\ [_{Path}\ TO\ ([_{Place}\ IN\ ([_{Thing}\ ]_j\ )])])]$

As it is possible to observe in the lexical entry of *enter* in (31)b (Jackendoff 1990:46), the difference between manner of motion verbs such as *run* and directed motion verbs such as *enter* is that *enter* incorporates the specification of the path argument. The sentence *John ran into the room* results, thus, from the combination of the LCS of the verb and the LCS of the PP *into*. The use of conceptual primitives to represent the semantic and syntactically relevant aspects of meaning (for instance GO, TO and IN), and constants, which capture the idiosyncratic elements of the meaning, allow the differentiation of semantically close verbs.

However, ultimately, this may result in different lexical entries for manner of motion verbs such as *run*, to account for the different syntactic constructions that are mirrored in sentences such as *John ran* and *John ran into the room*, and also for the different aspectual values associated to them, since the syntactic structure is determined by the conceptual structure, and in the sentence *John run* there is no path established. Also, this approach does not account for the fact that the co-occurrence with source and goal denoting PPs is a recurrent construction that seems to affect not only the manner of motion verbs denoting activities but a larger set of activity denoting verbs.

On a different approach (Goldberg (1995), Gawron (1985, 1986), Pustejovsky (1991), Hoekstra (1992), Zubizarreta & Oh (2004)), the multiplication of lexical entries is avoided since it is considered that the accomplishment sense of sentences with directional expressions is achieved compositionally, through the interaction between the directional PPs meaning and the meaning of the base activity denoting verb. Thus, it is not necessary to consider two different lexical entries for verbs such as *run*, although it is necessary to account for directional expressions in the lexicon (see Zubizarreta & Oh (2004:26), chapter 1).

Portuguese manner of motion verbs do not seem to reflect different Aktionsart properties when occurring with directional expressions. The sentences with verbs such as *andar* (walk) (denoting an activity type event) (in (32)a) do not have a perfect accomplishment value (in (32)b).

(32) a. O João andou durante 2 horas/*em 2 horas.
(John walked for 2 hours/in 2 hours)

b. O João andou até ao parque *durante 2 horas/#em 2 horas.
(John walked to the park for 2 hours /in 2 hours)

This phenomenon can be interpreted as further evidence for considering directional structures: we have an activity denoting verb occurring in an accomplishment denoting construction. The result is a sentence that has an accomplishment interpretation (inhibiting the possibility of combining it with the phrase *for 2 hours*) but that still maintains some of the activity value denoted by the verb (rendering odd the possibility of combining it with the phrase *in 2 hours*). The directional structure does not proceed from a different verb with different Aktionsart values and argument structure: it adds an endpoint to the event denoted by the verb resulting in an accomplishment-denoting sentence.

For these reasons, it seems reasonable to consider compositionality in these cases, since the Aktionsart properties of verbs are not altered and the manner of motion sense of these particular verbs is maintained. Also, the fact that several activity denoting verbs that do not encode motion in their meaning occur with these expressions, being interpreted as directed motion events – more frequently in English but also in Portuguese, see (33) – constitutes strong evidence in favor of this thesis.

(33) a. Ela fumou continuamente de Londres até Paris.
(She smoked continuously from London to Paris.)

> "They [verbs that do not encode motion in their meaning] are the strongest piece of evidence to question the view that syntactic structure of directed motion sentences is a projection of the lexical properties of the verbs, since it would be implausible to defend a lexical entry involving motion for these verbs."
>
> Gutierrez (2001: 58)

It is possible to observe different behaviors concerning Portuguese verbs of movement occurring in these structures. There are verbs that can occur simultaneously with different PPs denoting source, goal and path, others only occurring with one of these phrases or with different PP combinations:

(34) a. O vento levou o barco <u>do cais</u>$_{SOURCE}$ <u>para o mar</u>$_{GOAL}$ <u>pelo canal</u>$_{PATH}$.
(The wind took the boat <u>from the pier</u>$_{SOURCE}$ <u>to the sea</u>$_{GOAL}$ <u>through the channel</u>$_{PATH}$)

b. O vento levou o barco <u>do cais</u>$_{SOURCE}$.
(The wind took the boat <u>from the pier</u>$_{SOURCE}$)

c. O vento levou o barco <u>para o mar</u>$_{GOAL}$.
(The wind took the boat <u>to the sea</u>$_{GOAL}$)

d. O vento levou o barco <u>pelo canal</u>$_{PATH}$.
(The wind took the boat <u>through the channel</u>$_{PATH}$)

e. O vento levou o barco <u>do cais</u>$_{SOURCE}$ <u>para o mar</u>$_{GOAL}$.
(The wind took the boat <u>from the pier</u>$_{SOURCE}$ <u>to the sea</u>$_{GOAL}$)

f. O vento levou o barco do cais$_{\text{SOURCE}}$ pelo canal$_{\text{PATH}}$.
(The wind took the boat from the pier$_{\text{SOURCE}}$ through the channel$_{\text{PATH}}$)

g. O vento levou o barco para o mar$_{\text{GOAL}}$ pelo canal$_{\text{PATH}}$.
(The wind took the boat to the sea$_{\text{GOAL}}$ through the channel$_{\text{PATH}}$)

There are verbs that occur with directional expressions but do not necessarily allow the co-occurrence with some of the PPs separately, and vice-versa:

(35) a. O cão moveu-se/correu do parque$_{\text{SOURCE}}$ para a casa$_{\text{GOAL}}$ pela estrada$_{\text{PATH}}$.
(The dog moved/ran from the park$_{\text{SOURCE}}$ to the house$_{\text{GOAL}}$ through the road$_{\text{PATH}}$)

b. ?*O cão moveu-se/correu do parque$_{\text{SOURCE}}$.
(The dog moved/ran from the park$_{\text{SOURCE}}$)

c. Ele escalou o Everest pelo trilho mais fácil$_{\text{PATH}}$.
(He climbed the Everest through the easiest trail$_{\text{PATH}}$)

d. ?*Ele escalou o Everest do sopé$_{\text{SOURCE}}$ para o cume mais alto$_{\text{GOAL}}$ pelo trilho$_{\text{PATH}}$.
(He climbed the Everest from the bottom$_{\text{SOURCE}}$ to the highest summit$_{\text{GOAL}}$ through the trail$_{\text{PATH}}$)

Naturally, the verbs selecting source or goal denoting arguments have some restrictions in what concerns the occurrence with directional expressions. Namely, some incompatibilities seem to be related to the homograph prepositions that introduce the arguments of the verbs and the directional prepositions, see (36)d.

(36) a. ?*O gato aproximou-se[8].
'the cat came closer-SE'

b. O gato aproximou-se do quarto$_{\text{GOAL}}$.
'the cat came closer-SE to.the bedroom$_{\text{GOAL}}$'

c. O gato aproximou-se do quarto$_{\text{GOAL}}$ pelo corredor$_{\text{PATH}}$.
'The cat came closer-SE to.the bedroom$_{\text{GOAL}}$ through.the corridor$_{\text{PATH}}$'

d. *O gato aproximou-se da cozinha$_{\text{SOURCE}}$ para o quarto$_{\text{GOAL}}$ pelo corredor$_{\text{PATH}}$.
'The cat came closer-SE from.the kitchen$_{\text{GOAL}}$ to.the bedroom$_{\text{GOAL}}$ through.the corridor$_{\text{PATH}}$'

The diverse behavior and co-occurrence restrictions among the members of the class of verbs of movement constitute interesting issues to be addressed. Desirably, the goal is to identify the semantic and syntactic properties responsible for this diversity and the way to state them in the lexical entries of verbs.

[8] Note that this sentence is well-formed if the goal location is recovered from context.

1.2.2.2.3.2 Middle construction

The middle construction can be characterized by its generic use (Ruwet 1972), but also by the following properties (described in Zwart (1998), Fong *et al.* (2000) and Kageyama (2002)):

(37) The middle construction

i) does not occur with adverbs that specify the speech point;

ii) does not allow conditional interpretation in absolute constructions;

iii) does not occur with expressed subjects (although necessarily understood).

According to several authors (Jaeggli (1986), Tenny (1987), Hoekstra & Roberts (1993), Fagan (1992)), this construction, is only possible if the logical object is "affected" (i.e. if the logical object is somewhat affected by the event denoted by the verb; within thematic role theory, receives a Patient thematic role). However, this assumption is not universally accepted (Fong *et al.* 2000) and does not apply to Romance languages (see Ruwet (1972), Zribi-Hertz (1987), Cinque (1988), Cornips & Hulk (1999), Hulk & Cornips (2000)), Portuguese included.

(38) a. Short story books read easily.

b. Livros de histórias lêem-se facilmente.
(Books of stories read-SE easily)

Typically, the middle construction in Portuguese occurs with the clitic *se*, resulting in a very similar construction to the passive with *se*. Nevertheless, there are several characteristic properties that allow us to distinguish it from the passive with -*se*, here listed in Table 1.

Middle construction	**Passive with -SE**
- the syntactic subject usually occurs in pre-verbal position	- the syntactic subject usually occurs in post verbal position
- the verb has to be in the present tense	- the verb does not have to be in the present tense
- requires an adverbial modifier	- does not require an adverbial modifier
- the syntactic subject has generic value, interpreted as "this type of..."	- the syntactic subject does not have generic value
- ascribes a generic state or property	- does not ascribe a generic state or property

Table 1: Middle construction vs. Passive with -se

Verbs of movement in Portuguese enter the middle construction. Note, however, that some verbs in the middle construction do not necessarily require the clitic *se*. It is also interesting to observe that this 'optionality' does not seem to be available for all verbs, especially for verbs with non-affected objects, as *ver* (see) in (39)d.

(39) a. Estas portas abrem(-se) facilmente. (These doors open easily)

b. Estas caixas recuam(-se) facilmente. (These boxes move back easily)

c. *Estas caixas empurram facilmente. (These boxes push easily)

d. *Modelos bonitos vêem com agrado. (Good looking models see with pleasure)

As a final remark, note that neither the verbs of Putting in Spatial Configuration (Levin 1993:112) nor the verbs of Putting with a Specified Direction (Levin 1993:114) occur in middle constructions in English, although in Portuguese they do. In English these verbs are often polysemic between a stative sense, (e.g., *The statue lies on the table*), and a "dynamic" or change of state sense (e.g., *They laid the statue on the table*), whereas in Portuguese stative senses are not available, which may condition their reading in the appropriate sense (cf. Levin 1993: 113).

(40) Livros altos inclinam-se facilmente.
(*Tall books lean easily)

Once again, it seems interesting to analyze this diversity and, desirably, to correlate it with semantic and/or syntactic properties that can be expressed in the lexicon.

1.1.2.2.3.3 Non-causative construction

Non-causative constructions are quite productive in Portuguese. According to Levin (1993: 30), they occur with verbs whose logical object is affected by the event denoted by the verb. Typically, this construction is licensed by change of state verbs and change of position and location verbs. In non-causative constructions, the logical object occurs in syntactic subject position, describing a process in which the mention of the cause of the event is taken as irrelevant, although it may appear in a PP introduced by the preposition *com* (≅ with). Usually, in Portuguese, non-causative construction does not require the presence of the clitic *se*, as demonstrated in (41)b. However, several verbs allow the absence of the clitic, in (41)a, and others require its presence, (41)c. The distribution of the clitic *se* in this construction, as well as in the middle construction, as seen above, is not, however, easily explained.

(41) a. A porta abriu-se/abriu (com o vento). (The door opened (with the wind))

b. O balão *subiu-se/subiu (com o vento). (The air balloon ascended (with the wind))

c. O balde moveu-se/*moveu (com o vento). (The bucket moved (with the wind))

Other differences between English and Portuguese verbs can be observed in verbs of Putting with a Specified Direction (Levin 1993:114), which in English do not allow causative/non-causative alternations whereas in Portuguese they do.

(42) a. O vento levantou a tampa do alçapão.
(The wind lifted the pitfall door)

b. A tampa do alçapão levantou/levantou-se com o vento.
(*The pitfall door lifted wind the wind)

It is our perspective that some of these differences may be related to the semantic properties of particular verbs, although not common to the entire class. The determination of such properties, their representation in lexical entries and their correlation to the syntactic constructions allowed constitutes, as stated before, one of goals of the current work.

1.3 Research directions

Given our starting point – the verbal lexicon and, more specifically, the class of verbs of movement in Portuguese – and the goals in pursue – the construction of a verbal lexicon that, besides modeling linguistic knowledge, reflecting concerns regarding complexity, considers the representation of information useful for computational purposes –, the development of the work depicted in this dissertation has to consider two main issues: the design and organization of the lexicon and the establishment and modeling of the relevant information in the lexical entries.

One of the essential steps for the construction of a model for the computation of meaning is the structure of the Lexicon: the way lexical items are organized can and should contemplate the conceptual, semantic and syntactic properties of lexical items, as well as regular processes of lexicalization (Kilgarrif 1993). For these reasons, lexical resources have to consider the organization of verbs also according to the network of relations established between them, and the definition of a lexical inheritance device that allows the sharing of the relevant syntactic and semantic information.

On the other hand, the information at the lexical level should consider the polymorphic properties of natural language and the creative use of words, specifically their context sensibility (see Pustejovsky (1993, 1995), Fellbaum (1998a, 1998b, 1999), Buitelaar (1998), Fong *et al.* (2000), among others). This leads us to the need of considering lexical units as informational structures, which, along with computational motivations is also psychologically motivated (see Sag & Wasow (1999)).

1.3.1 WordNet model

For the organization of the verbal lexicon in study under the scope of our work, we adopted a relational model of the lexicon, following the Princeton WordNet (Miller *et al.* (1990), Fellbaum (1998c)) and EuroWordNet models (Vossen 2003). Besides being lexical resources with numerous applications in the domain of computational linguistics (see Hanks (2003)), wordnets also reflect the organization of the mental lexicon and have been proved of great psychological realism (see all the research work referred in http://www.globalwordnet.org).

A wordnet is an electronic relational lexical database that combines a thesaurus with an ontological database, in which the lexical items are organized in sets of synonyms representing concepts (synsets), related to each other through several types of relations: lexical relations (synonymy), lexical-conceptual relations (hyperonymy/hyponymy, meronymy, etc.), function or role relations (agent_patient relation, involved_instrument relation, etc.), semantic opposition relations (antonymy, near_antonymy) and cause relations (is_caused_by/causes, etc.) (see Marrafa (2001, 2002)). The structuring relation in wornets is the hyponymy/hyperonymy relation. In general terms, hyponymy relation is defined as follows:

(43) X is hyponym of Y, if
X is a type of Y and Y is not a type of X.

Hyponymy/hyperonymy is a lexical-conceptual relation that concerns both world-knowledge and linguistic knowledge, as showed by anaphoric constructions such as the ones presented in (44), where the hyperonym is used to refer a more specific referent (the hyponym) previously introduced.

(44) a. He bought a <u>German Shepard</u> but the <u>dog</u> doesn't bite.

b. He <u>crawled</u> through the woods, <u>moving so</u> to avoid being seen by the guards.

c. <u>Marine</u> animals can be endangered by the near proximity of cities, since many <u>aquatic</u> animals are easily affected by sewage pollution.

The hierarchical nature of the hyponymy relation can be tested in contrastive contexts and simple coordination structures, which show the meaning differences between the hyponym and hyperonym:

(45) a. #He bought a <u>German Shepard</u> and a <u>dog</u>.

b. #A <u>German Shepard</u> is more aggressive than a <u>dog</u>.

c. #He owned a <u>German Shepard</u>, but not a <u>dog</u>.

c′. He owned a dog, but not a German Shepard.

d. #He crawled and moved through the woods to avoid being seen by the guards.

e. #He crawled through the woods, but he did not move.
e′. He moved through the woods, but he did not crawl.

f. #He saw marine and aquatic animals.

g. #He saw marine but not aquatic animals.
g′. He saw aquatic but not marine animals.

This way, lexical items are organized according to their *type*, a hyponym being all that its hyperonym is, and more. This relation contemplates at the same time the definition of a monotonic inheritance device (see Miller 1990) that allows an adequate description of lexical items in an economic way, since hyponyms naturally inherit the conceptual properties of their hyperonym.

Verbal hyperonymy relation is defined in slightly different terms than that of nouns. Verbal hyponymy, also dubbed *troponymy* by Fellbaum (1998a), is defined as follows:

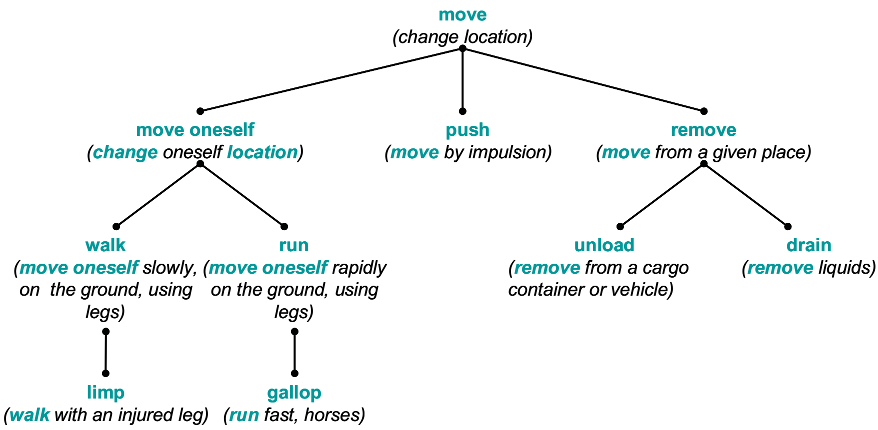
(46) a. V1 is a troponym of V2, if
to V1 is to V2 in some particular manner. (Fellbaum 1998a:79).

b. *crawl* is a troponym of *move*, if
to *crawl* is to *move* in some particular manner (*with the body near to the ground)*.

A relational analysis of the lexicon represents also an analysis in terms of semantic domains, since, as stated in Fellbaum (1998a), certain types of lexical and semantic relations usually occur within the same semantic domain. Also, if we consider that the verb hyperonymy relations encode the core meaning of a given verb (corresponding to the meaning of the hyperonym), the determination of a given class of verbs can be quite easily achieved and lexical-conceptually motivated, since all the hyponyms of a given verb are surely of the same semantic class.

Considering the class of verbs of movement, object of the current work, it is reasonable to assume that all its members will be lexical-conceptually related and therefore that this representation model will serve the goals in pursue in a cost-effective way.

(47) Example of verbs of movement organized in a relational net



Fellbaum (1998a) also notes that it is possible to observe several regularities in the English relational net for verbs, regarding the hyponymy level depth and the number of hyponyms in each level. For instance: the deeper the level of hyponymy, the greater the number of hyponyms and the more specific the concepts denoted. Consequently the stricter the selection constraints imposed by the hyponym verbs. Aspects such as these reveal the ability this model has to reflect lexicalization patterns, but also constitute evidence enabling the determination of the semantic elements that are specific to each verb, and the consequent implications for the computation of its meaning (see Krifka (2001), Fong & Fellbaum (2003), among others).

Thus, in our perspective, the lexical-conceptual organization has several advantages: it assures semantic grounds for establishing classes, it provides evidence for similar semantic (and desirably syntactic) behavior among lexical items and, more importantly, it can guide us in the identification of the relevant information representing the meaning of verbs, which can help us to accurately predict their syntactic behavior, co-occurrence restrictions and new meanings in context.

1.3.2 Generative Lexicon model

The goals of the present work, combined with the nature its object and the range of phenomena considered, induce the need to consider lexical units as informational structures. In the wordnet model, the semantic information that characterizes a given item is not overtly represented, since the meaning of a given item is deductible from its position in the net, i.e., from the lexical-conceptual relations that establish its position. For these reasons, we propose the modeling of the linguistic information to be stated in lexical entries within the Generative Lexicon (GL) framework (Pustejovsky 1991, 1995).

In GL, the Lexicon is viewed as a complex system that allows generativity. The representation of lexical units as informational structures, according to a finite set of rules, allows the description of meaning in context and of the relation between semantics and syntax. The information is represented in a well-defined feature structure system, in attribute-value matrixes (AVM), whose values can be atomic or, recursively, other AVMs. The systematic and declarative way in which the semantic content of lexical units is represented assures the coherent and recursive codification of the information, but also enables the sharing of information between structures, resulting in a relatively simple but adequate modeling. Also, the operational devices defined in this model – the generative mechanisms of type-coercion, co-composition and selective binding – allow the use of the lexical representations compositionally.

In the GL model, lexical entries are structured according to four levels of representation: event structure, argument structure, qualia structure and lexical inheritance structure. The study of verbal predicates, namely of those involving alternation constructions, shows that the explanation of divergent semantic and syntactic behaviors has to consider the internal structure of the predicates – the event structure – and argument selection properties – the argument structure.

The event structure contemplates the logical representation of the internal structure of a given predicate, including the listing of events associated to it, partial order constraints, overlapping, inclusion and definition of the event head. The argument structure contemplates the definition of the semantic properties of the prototypical arguments of the predicates as well as syntactic mapping information. In what concerns verbal predicates, "event structure can provide a distinct and useful level of representation for linguistic analysis involving the aspectual properties of verbs, adverbial scope, the role of argument structure, and the mapping from lexicon to syntax" (Pustejovsky 1991:47).

The qualia structure can be briefly described as "that set of properties or events associated with a lexical item which best explain what that word means" (Pustejovsky 1995:85). The direct or indirect reference to qualia information in the lexical entries has been consistent within the Lexical Semantics field and its integration in relational models of the lexicon – for co-hyponyms differentiation purposes, for instance – has been studied by several authors (see Busa *et al.* (2001), Mendes & Chavez (2001), Vossen (2001), Marrafa (2002), O'Hara & Wiebe (2003), Veale (2003), Pederson & Sorensen (2006), among others).

The lexical inheritance structure is the level of representation that concerns the percolation of information through the lexical entries, to avoid the repetition of information on one hand, but also to regulate the information to be shared. In GL, this structure identifies "how a lexical structure is related to other structures in the type lattice, and its contribution to the global organization of a lexicon" (Pustejovsky 1995: 61). This way, it is possible to define the inherited semantic features of a given structure, through the connection to a given semantic type, inserted in an orthogonal type lattice.

The levels of representation and the generative mechanisms considered in GL, and the range of phenomena this particular framework aims to cover, constitute the grounds in which we base our choice. Also, already developed research concerning the association of GL structures to relational models of the lexicon seems to validate the assumption that wordnets can constitute the base for generative lexica, since it is possible to integrate additional information to a relational model such as WordNet without compromising its global structure (Amaro (2006), Amaro *et al*. (2006)).

Nonetheless, there are still several issues to be addressed concerning different approaches to regular phenomena, such as verbal alternations. For the sake of reasoning, consider the example sentences in (49):

(48) a. The wave sunk the boat.

b. The boat sunk.

According to Pustejovsky (1991, 1995), the information in the lexical entry of the verb *sink*, namely the fact of it being a verb denoting a complex event (consisting of a process event that precedes a final state event) with no head event defined, accounts for the alternation constructions:

(49) a. *sink* = e_1: process < e_2: state, head =_.

b. The wave sunk the boat.
(*sink* = e_1: process < e_2: state, head = e_1)

c. The boat sunk.
(*sink* = e_1: process < e_2: state, head = e_2)

In (50)b, the prominent event is the process (e_1) that leads to the final state (e_2), resulting in a causative reading. In (50)c, the prominent event is the final state (e_2), resulting thus in an inchoative reading. In GL, a single lexical entry can account for the causative/inchoative alternation, given that it considers the contribution of the linguistic context in which a given lexical item occurs in the determination of a given reading, i.e., meaning in context.

In the Princeton WordNet model (Fellbaum 1998c), this phenomenon is addressed by the statement of two distinct lexical entries: $\{sink_1\}_{Vcausative}$ and $\{sink_2\}_{Vinchoative}$, connected by a cause relation in which $\{sink_1\}_V$ presupposes but does not include $\{sink_2\}_V$. In WordNet.PT (Marrafa 2001, 2002) this strategy is avoided, since it is argued that in both causative and inchoative sentences the event denoted by the verb is the same, although the alternative constructions reflect the different prominence given to the subevents (see Marrafa & Mendes (2007)).

Note, however, that the analysis of these issues further motivates the need for richer structures at lexical entry level.

Also, and naturally, the issues directly related to the articulation of the GL semantic and syntactic lexical representation with the configuration of the WordNet model constitute also a substantial part of this work.

1.4 Path

The organization of this dissertation mirrors, to some extent, the steps taken to complete the work it depicts. Considering the object and the goals we present in this chapter, we proceed, in chapter 2, to the presentation of the Portuguese wordnet of verbs of movement developed, articulated with the characterization of the WordNet model, with particular attention to the specificities of the Portuguese WordNet (WN.PT). We also present the analysis of Portuguese verbs of movement in what concerns the definition of the primary nodes, the structure of the relational net and the definition of additional organizational criteria.

In chapter 3, we focus on the lexicalization of semantic components in Portuguese verbs of movement and on hyponymy analysis, in order to identify the semantic elements that differentiate the nodes in the lexical-conceptual network. We present our first analysis of Portuguese data, in what concerns the lexicalization patterns encountered, co-hyponym compatibility (i.e. the possibility of co-occurrence of sister nodes), and the correlation between the lexicalization of semantic components at the hyponym level, verbal argument structure and Aktionsart properties.

Chapter 4 is dedicated to the presentation of the Generative Lexicon, in which we model argument, event and qualia information of the verbs in study. We focus on the representation levels and other elements of this framework, and on some adjustments necessary to a coherent modeling of all POS, proposing also how the assimilation of wordnet hierarchical structure in GL is possible.

To account for argument structure, and specifically for subcategorization properties of verbs, chapter 5 addresses the representation of prepositions. Profiting from already developed work, we model prepositions in a wordnet, and propose their semantic representation at lexical entry level within the GL model.

Chapter 6 is dedicated to the presentation of the resulting lexical structures, considering GL formulations discussed in chapter 4 and the lexical inheritance structure established through hyponymy. We address illustrative cases, the particular reflections of semantic lexicalization in the semantic and syntactic properties of hyponym verbs and the patterns that emerge from the set of verbs analyzed, and formalize verbal co-hyponym compatibility through qualia unification.

In chapter 7 we propose the integration of the GL levels of representation in WN.PT, namely argument structure, qualia structure and event structure, demonstrating how wordnets can support a finer-grained lexical description framework that provides the bases for accounting for several lexical semantic phenomena.

Chapter 8 is dedicated to the analysis of the constructions in which Portuguese verbs of movement occur and to the correlation of occurrence restrictions with the lexical semantic properties of these verbs. As briefly presented above, we will focus on directed motion constructions, middle constructions and non-causative constructions, but also on obstacle, region and path denoting arguments.

Finally, chapter 9 summarizes the conclusions depicted throughout the remainder of this dissertation, presenting some final remarks regarding the work developed and the achievement of the goals weset ourselves to attain.

2. WordNet.PT: encoding verbs of movement

2.0 Introduction

This chapter is dedicated to the construction of a fragment of Portuguese wordnet encoding verbs of movement. As it will be showed here, the process of building a relational lexicon contemplates different levels of analysis that are relevant for achieving our goals.

Briefly, the construction of a wordnet is made of small steps, not necessarily simple, and starts with collecting possible candidates (through several possibly different strategies: from our own lexical knowledge, through the inquiry of other native speakers, from lexical resources such as dictionaries, encyclopedias, treebanks, wordnets in other languages, etc.), usually within a given semantic domain. The main task concerns the organization of these items in synonym sets, according to the concept they denote, and in subtyping relations. Subtype information is frequently used in paraphrases or definitions of what a given word means. Thus, typically, these tasks are closely intertwined. The last step is to link a given synonym set to other concepts already established in the net, through the different relations available in the model.

The results necessarily reflect generalizations about the lexicalization of concepts of a given language but they also reflect the ability of the model to mirror the lexical-conceptual relations established between the concepts in the lexicon.

For this reason, we divided this chapter in two major subsections. Section 2.1 is dedicated to the characterization of the WordNet model (WN), in general, the EuroWordNet model (EWN), in particular, and the specificities of the Portuguese WordNet (WN.PT), in which the data were implemented. Section 2.2 presents our analysis of Portuguese verbs of movement regarding the definition of the primary nodes, the structure of the relational net and the definition of additional organizational criteria. Finally, section 2.3 concerns the conclusions to be drawn from the accomplishment of this task and the relevant issues requiring further analysis.

2.1 WordNet

It is fair to say that the WN model was not originally designed to be a traditional lexical resource, but evolved from the "need for a comprehensive lexical database that would include word meanings as well as word forms and that could be used under computer control." (Miller 1998: xvi). Originating from George Miller's research in the early 80's, WN profited from experiments on the organization of the mental lexicon, which pointed out its relational nature, confirming previous lexicography research on the semantic structure of dictionaries (Litkowski 1978), and being corroborated by subsequent research in several domains, such as lexicalist models of grammar (Bresnan 1982, Pollard & Sag 1994) or generative models of the lexicon (Pustejovsky 1995).

The WN model is currently recognized as a revolutionary reference system, which combines a thesaurus with an ontological database (see Hanks (2003), among others), and has been extensively adopted for the construction of robust lexical resources in many languages, used in several areas of Computational Linguistics and Language Engineering, namely for information retrieval, word sense disambiguation, question answering, text summarization, and semantic information extraction (see Litkowski (2005)).

2.1.1 WordNet 1.5: "the mother of all wordnets"

The basic semantic relation in WN is SYNONYMY. The nodes of the net are sets of synonym words of the same POS that correspond to the lexicalization of a concept in a given language. The SYNONYMY relation defined here does not correspond to absolute or true synonymy (by which two words are synonyms iff the replacement of one for the other does not change the truth value of any sentence in which they occur). Rather, it corresponds to synonymy in context, defined in Miller *et al.* (1990) as follows:

(1) "… two expressions are synonymous in a linguistic context C if the substitution of one for the other in C does not alter the truth value" (Miller *et al.*, 1990:6).

According to this definition, SYNONYMY necessarily considers the semantic domain in which the concept denoted by the word is inscribed. Take, for instance, the following examples of verbs in WordNet 1.5 (WN 1.5):

(2) a. John rose the staircase.

b. John ascended the staircase.

rise and *ascend* are synonyms in this semantic domain.

(3) a. John rose to the challenge.

b. *John ascended to the challenge.

rise and *ascend* are not synonyms in this semantic domain.

The SYNONYMY relation is necessarily a symmetric relation.

(4) SYNONYMY relation

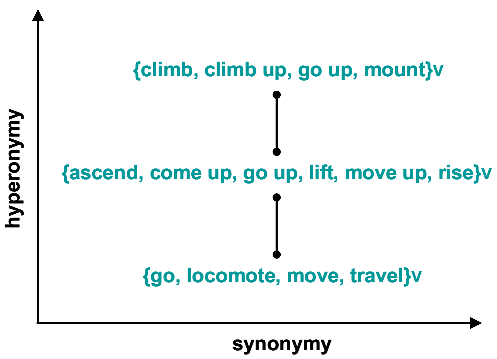
a. if *A* and *B* are synonyms, in a context C, then
   (i) *A* is *B* in C, and
   (ii) *B* is *A* in C.

(5) a. if *rise* and *ascend* are synonyms, in a context C, then
   (i) *rise* is *ascend* in C, and
   (ii) *ascend* is *rise* in C.

The SYNONYMY relation adopted in WN illustrates the first structural difference opposing it to more conventional dictionaries: in WN units do not correspond to a list of possible meanings of a given word; instead, a word is part of as many units (synonym sets: synsets) as the concepts it denotes. Synsets (sets of synonym words lexicalizing concepts), rather than words, are the basic units of WN and SYNONYMY can, thus, be considered the horizontally structuring semantic relation of the model.

Also regarding the structure, the other structuring relation is the subtyping relation, "subordination (or class inclusion or subsumption)" (Miller 1998: 24) holding between lexicalized concepts. The HYPERONYMY/HYPONYMY relation (in WN terminology) is the semantic relation that allows the establishment of a lexical hierarchy in the lexicon, corresponding to the vertically structuring relation of the model.

(6) WN 1.5 excerpt



As briefly stated in the previous section, verbal HYPONYMY, dubbed TROPONYMY (Fellbaum & Miller 1990, Fellbaum 1998a), does not correspond exactly to nominal HYPONYMY.

(7) Definition of nominal HYPONYMY in WN 1.5
X is hyponym of Y, if
X is a kind of Y. (Miller 1998:34)

Verbal HYPONYMY, in fact, is not exactly a subtyping relation, but rather a particular kind of entailment. This distinction is motivated by three major reasons:

i) Speakers are not keen in accepting the hyponymy paraphrase applied to verbs (*climb is a kind of ascend*), nor the omission of the "a kind of" from the formula, acceptable and common for nouns: *A horse is an animal* vs. *To climb is to ascend* (but rather *To climb is to ascend in some particular manner*).

ii) Hyponym verbs entail their hyperonym. *Climb* entails *ascend*: when someone is climbing then someone is ascending (but *ascend* does not entail *climb*: when someone is ascending, someone is not necessarily climbing).

iii) The events denoted by a verbal hyponym and by its hyperonym are always temporally coextensive. Once again, someone is necessarily ascending for as long as someone is climbing (but someone is not necessarily climbing for as long as someone is ascending).

In this way, in WN 1.5 terms, verbal HYPONYMY, can be defined as follows:

(8) V_1 is troponym of V_2, if
to V_1 is to V_2 in some particular manner, and
V_1 entails V_2. (Fellbaum 1998a: 79, 80)

ENTAILMENT, without full temporal co-extensiveness, can also link verbal concepts in WN 1.5, and constitutes the lexical relation correspondent to MERONYMY (is part of) for nouns. For instance, verbs such as *sleep* and *snore* or *forget* and *know* are related to one another by ENTAILMENT but with different temporal inclusion relations:

(9) *snore* entails *sleep* (when someone is snoring then someone is sleeping), and *sleep* properly includes *snore* (someone is necessarily sleeping when someone is snoring). (Fellbaum 1998a: 80, 84)

(10) *forget* entails *know* (when someone forgets then someone knew), and *forget* backward presupposes (i.e. does not temporally include) *know* (when someone forgets then someone had to previously know).(Fellbaum 1998a: 82, 84)

Verbs can also be linked by a CAUSE relation, being verbs in causative/inchoative alternations or not. In the first case, we have verbs such as *break*$_{[\text{causative}]}$ - *break*$_{[\text{inchoative}]}$ (*to break a chair* causes *the chair to break* (better still, to be broken), and *the chair breaking* (being broken) is necessarily caused by *break*). In the second case, we have verbs such as *show-see*, *kill-die* (*show* causes *see*, but *see* is not necessarily caused by *show*; *kill* causes *die*, but *die* is not necessarily caused by *kill*).

Semantic opposition is also expressed in WN 1.5, either by OPPOSITION pointers, that express "a complex relation encompassing several distinct subtypes of semantic opposition" (Fellbaum 1998a: 81), or by the lexical relation of ANTONYMY.

OPPOSITION pointers establish a relation between synsets, i.e., between concepts, and can link converse verbs, such as *give/take*, *buy/sell*, *lend/borrow*, *teach/learn* (Fellbaum 1998a: 81); opposite but not antonymous verbs, such as *fall/ascend*; co-hyponym verbs with opposite or contrasting particular manners (given their hyperonym) such as *walk/run*; and verbs related by backward presupposition (i.e. the event denoted by V1 is entailed by and precedes the event denoted by V2), such as *know/forget*, *tie/untie*, *wrap/unwrap*.

ANTONYMY is the relation that holds between two given word forms that lexicalize contrary concepts, which can be typically expressed in terms of attributes. For instance, although *fall* and *ascend* are opposite verbs, *fall* and *rise*, on the one hand, and *ascend* and *descend*, on the other, are antonym words, but *fall* and *ascend* are not.

Relations linking synsets implemented in WN 1.5 were intentionally small in number, and can be summarized in the following table of pointers for verbs and nouns (adapted from Tengi (1998: 109)):

	Verbs	**Nouns**
Pointers for lexical relations	antonym – antonym	antonym – antonym
Pointers for lexical-conceptual relations	hyperonym – troponym	hyperonym – hyponym
	entails – is entailed	holonym – meronym
	causes – is caused by	

Table 1: Pointers indicating the major relations for Verbs and Nouns in WN 1.5

WN 1.5 considers the four major POS – nouns, verbs, adjectives and adverbs – and defines some cross-POS relations, typically related to morphological features, such as *derived from* (between adverbs and adjectives) and *participle* (between adjectives and verbs), or semantic relations such as *attribute* (between nouns and adjectives), and particular POS relations such as *similar to* and *relational adjective* (for adjectives) and *also see* (for verbs). However, given the separate database files in which the different POS are stored, these were not originally implemented.

A great deal of information is also directly or indirectly stated in synset definitions or glosses. Although not entirely consistent with the theoretical grounds of the model – according to which the meaning of a given synset (and naturally of the specific words that compose it) is established by its position within the net (hyperonymy/hyponymy relation) defined by the relations established with the other nodes in the net – glosses or definitions allow for a more user friendly consultation of the lexical database by the human users. Glosses or definitions are associated to a given synset and can include simple explanatory sentences or usage examples.

(11) Information on the synset {go, locomote, move, travel}$_V$ in WN 1.5

{go, locomote, move, travel [change location; move, travel, or proceed; "How fast does your new car go?" "We traveled from Rome to Naples by bus"; "The Policemen went from door to door looking for the suspect"]}$_V$

In WN 1.5 verbal synsets are also associated to a list of verb frames that exemplify basic syntactic constructions in which the verbs of a given synset, or a specific verb within a given synset, can occur. The following list constitutes part of the 34 verb sentence frames considered in WN 1.5 (taken from Miller *et al.* (1990: 80))[1].

(12)

Appendix B
Verb Sentence Frames

1 Something ----s
2 Somebody ----s
3 It is ----ing
4 Something is ----ing PP
5 Something ----s something Adjective/Noun
6 Something ----s Adjective/Noun
7 Somebody ----s Adjective
8 Somebody ----s something
9 Somebody ----s somebody
10 Something ----s somebody
11 Something ----s something
12 Something ----s to somebody
…

WN model has been adapted to several languages worldwide and the relations and implementation design of the model have been modified and enhanced, reflecting research in several areas of Lexical Semantics.

2.1.2 EuroWordNet

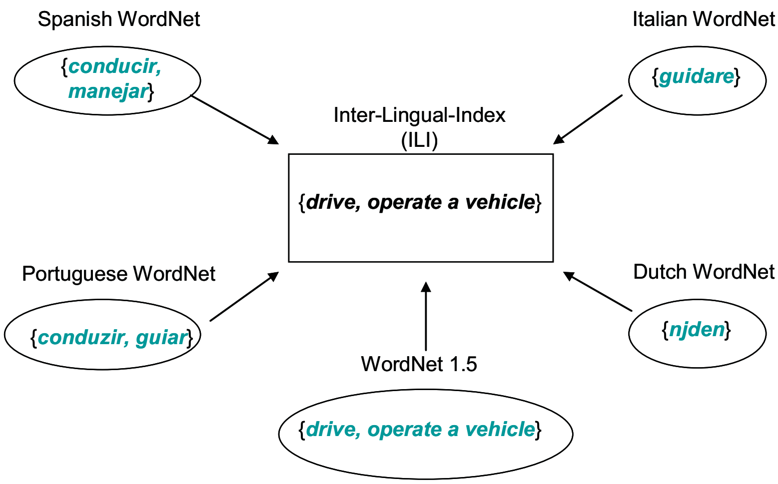
EuroWordNet (henceforth EWN) is probably one of the major developments in European lexical resources that embraced WN 1.5 as its base model, constituting itself real evidence of the plasticity and universal character of this model.

[1] For a more detailed description of WN 1.5 design and implementation, see Miller *et al.* (1990: 62-77).

Following the structural relations defined by WN 1.5, EWN differs from it for being a multilingual database (which integrates databases for several European languages, related to each other), for uniformly treating all POS within the same database file and for using a larger number of relations and labels.

EWN gathers wordnets for several European languages, developed individually according to the set of criteria established in Vossen (2002), and connected to each other through an Inter-Lingual-Index (ILI). The ILI is an index of synonym sets, with no internal structure, extracted from WN 1.5. Basically, it provides a common link that makes the mapping of the synsets in one particular language to correspondent synsets in the other languages. Although flat in structure, the ILI is also connected to WN 1.5, allowing the recovery of the English structure.

(13) EWN architecture (adapted from Marrafa (2001: 20))



The second difference between the EWN model and WN 1.5, which regards the treatment of all POS within the same database file, although seeming only an implementation option, constitutes a major factor in what concerns the relations available in the model. The stricter division of POS in WN 1.5 prevented cross-POS relations to be implemented, namely attributive features linking nouns and adjectives and functional features linking nouns and verbs; the latter, even though described and motivated, were not implemented in WN 1.5 (see Miller *et al.* (1990: 18-19)).

In EWN, although the nodes correspond to concepts necessarily lexicalized by lexical items of the same POS, distinctions are made in terms of type of entities (following Lyons (1977)) rather than in terms of POS. In EWN three types of entities are considered:

i) *First Order Entities*: concrete entities "perceivable by the senses and located at any point in time, in a three-dimensional space, e.g. object, substance, animal, plant" (Vossen 2002: 12);

ii) *Second Order Entities*: static and dynamic situations that "cannot be grasped, heard, seen, felt as an independent physical thing. They can be located in time and occur or take place rather than exist; e.g. be, happen, cause, move, continue, occur, apply." (Vossen 2002: 12);

iii) *Third Order Entities*: unobservable propositions that exist "independently of time and space. They can be asserted or denied, remembered or forgotten. E.g. idea, thought, information, theory, plan, intention." (Vossen 2002: 12).

This strategy, besides allowing cross-POS relations, does not limit a large number of relations to specific POS. For instance, a deverbal noun such as *destruction* can be related to other nouns by role relations, maintaining the argument structure of the event.

Briefly, the following tables summarize the relations available in EWN, the tests devised and some examples, adapted from Vossen (2002).

Relation	Tests	Examples
is synonym of	For nouns a) if it is a N_1 then it is a N_2; and if it is a N_2 then it is a N_1 For verbs a) if someone/it V_1s then someone/it V_2s; and if someone/it V_2s then someone/it V_1s	*fiddle* **is synonym of** *violin* *violin* **is synonym of** *fiddle.* *begin* **is synonym of** *start* *start* **is synonym of** *begin*
is antonym of	For nouns a) N_1 and N_2 are both a kind of N_3 but N_1 is the opposite of N_2 and N_2 is the opposite of N_1. For verbs a) V_1 and V_2 are both hyponym of V_3; b) if someone/it V_1s then someone/it does not V_2; and if someone/it V_2s then someone/it does not V_1.	*human being* has hyponyms *man* and *woman* *man* **is antonym of** *man* *woman* **is antonym of** *man* *go* has hyponyms *enter* and *exit* *enter* **is antonym of** *exit* *exit* **is antonym of** *enter*
derives from/has derived	*Morphological relation expressing derivational processes. Typically holds between two semantic related variants.*	*richness* **derives from** *rich* *rich* **has derived** *richness*
pertains/ is pertained	*Morphological relation expressing a somewhat unclear semantic relation. Typically holds between adjectives and nouns to which the first are related.*	*atomic* **pertains to** *atom* *atom* **is pertained** *atomic*

Table 2: EWN relations available between word forms

Relation	Tests	Examples
has hyponym/has hyperonym	For nouns a) a N_1 is a N_2 with certain properties; OR b) a N_1 is a kind/type/species/brand of N_2. c) It is a N_1 and therefore also a N_2. (True) It is a N_2 and therefore also a N_1. (False) For verbs a) to V_1 is to V_2+AdvP$_i$/AdjP$_j$/NP$_k$/PP$_l$; but to V_2 is not to V_1+AdvP$_i$/AdjP$_j$/NP$_k$/PP$_l$	{car}$_N$ **has hyperonym** {vehicle}$_N$ {vehicle}$_N$ **has hyponym** {car}$_N$ {dog}$_N$ **has hyperonym** {animal}$_N$ {animal}$_N$ **has hyponym** {dog}$_N$ {run}$_V$ **has hyperonym** {go}$_V$ {go}$_V$ **has hyponym** {run}$_V$
is meronym of/ is holonym of (N)	a) a N_1 makes up a part of a N_2 and a N_2 has N_1s, but not the converse. b) a N_1 is a member/element of a N_2 and a N_2 has N_1s, but not the converse. c) a N_1 is an amount/piece/portion of N_2 and a N_2 has N_1s, but not the converse. d) a N_1 is a component of a N_2 and a N_2 has N_1s, but not the converse. e) a N_1 is made of N_2 and N_2 is a substance, but not the converse. f) a N_1 is a place located in N_2 and N_2 has N_1s, but not the converse.	{skin}$_N$ **has holonym** {body}$_N$ {body}$_N$ **has meronym** {skin}$_N$ {player}$_N$ **has holo_member** {team}$_N$ {team}$_N$ **has mero_member** {player}$_N$ {drop}$_N$ **has holo_portion** {liquid}$_N$ {liquid}$_N$ **has mero_portion** {drop}$_N$ {wheel}$_N$ **has holo_part** {car}$_N$ {car}$_N$ **has mero_part** {wheel}$_N$ {wood}$_N$ **has holo_madeof** {stick}$_N$ {sitck}$_N$ **has mero_madeof** {wood}$_N$ {centre}$_N$ **has holo_location** {city}$_N$ {city}$_N$ **has mero_location** {centre}$_N$
is near synonym of	N_1 is hyponym of N_3 and N_2 is hyponym of N_4 a) if it is a N_1 then it is also a kind of N_4 but you usually do not call N_1 an N_4, and if it is a N_2 then it is also a kind of N_3 but you usually do not call N_2 an N_3.	{hammer}$_N$ **is hyponym of** {tool}$_N$ {measuring rod}$_N$ **is hyponym of** {instrument}$_N$ {tool}$_N$ **is near synonym of** {instrument}$_N$ {instrument}$_N$ **is near synonym of** {tool}$_N$

Table 3: EWN relations available between synsets of the same POS

All the relations in tables 2 and 3 are also available in WN 1.5, exception being the NEAR SYNONYMY relation, which expresses semantic proximity and almost synonymy between two synsets.

However, it is possible to see several specifications of the MERONYMY relation, which in EWN is further divided according to the type of part-whole relation established, which depend on the entities linked this way: a given entiey is composed of a set of discrete entities (HAS_MERO_MEMBER/HAS_HOLO_MEMBER); a pre-existing substance is typically divided in undefined (but lexicalized) portions (HAS_MERO_PORTION/HAS_HOLO_PORTION); a concrete entity is composed of several concrete entities with definite bounderies and functions (HAS_MERO_PART/HAS_HOLO_PART); a concrete entity is composed of a given substance/matter

(HAS_MERO_MADE_OF/HAS_HOLO_MADE_OF); and a given location includes a lexicalized spatial subdivision (HAS_MERO_LOCATION/HAS_HOLO_LOCATION).

It is also interesting to notice that the tests for HAS HYPERONYM/HAS HYPONYM relation focus on nouns and verbs, reflecting language major denoting POS, but also mirroring the specificities of the lexical-conceptual organization of other POS considered in EWN – adjectives and adverbs –, which exhibit shallow or even inexistent hierarchies of concepts.

The next table presents the rest of the relations considered in EWN.

Relation	Tests	Examples
is subevent of/ has subevent (N and V)	a) N_1/V_1 takes place during or as part of N_2/V_2, and whenever N_1/V_1 takes place, N_2/V_2 takes place, but not the converse. b) N_2/V_2 consists of N_1/V_1 and other events or processes, but not the converse.	{dream}V **is subevent of** {sleep}V {sleep}V **has subevent** {dream}V *reversed* {oxigenation}N **is subevent of** {breathing}N {breathing}N **has subevent** {oxigenation}N {analysis}N **is subevent of** {research}N {research}N **has subevent** {analysis}N
causes/ is caused by (N, V and Adj)	a) to/a V_1/N_1 causes to/a V_2/N_2 to take place, OR to/a V_1/N_1 has to/a V_2/N_2 as a consequence, OR to/a V_1/N_1 leads to/a V_2/N_2, but not the converse. b) V_1/N_2 may cause V_2/N_2, OR V_1/N_1 may have V_2/N_2 as a consequence, OR V_1/N_1 may lead to V_2/N_2, but not the converse. c) V_1 causes to be Adj_1, OR V_1 has being Adj_1 as a consequence, OR V_1 leads to being Adj_1, but not the converse. d) If V_1/N_1 takes place it causes/may cause V_2/N_2 to take place afterwards, but not the converse. e) V_1/N_1 is not hyperonym of V_2/N_2, and if V_1/N_1 takes place it causes/may cause V_2/N_2 to take place at the same time, but not the converse	{kill}V **causes** {die}V {die}V **is caused by** {kill}V *reversed* {kill}V **causes** {death}N {death}N **is caused by** {kill}V *reversed* {kill}V **causes** {dead}Adj {dead}Adj **is caused by** {kill}V *reversed* {search}V **causes** {find}V *non-factive* {find}V **is caused by** {search}V *non-factive* {pull}V **causes** {open}V *non-factive* {open}V **is caused by** {pull}V *non-factive*
is XPOS near synonym (N, V and Adj)	a) if there is a case of N_1 then someone/it V_1s, and if someone/it V_1s then there is a case of N_1. b) if a N_1 takes place then someone/it V_1s, and if someone/it V_1s then a N_1 takes place. c) if there is a state of N_1 then someone/it V_1s, and if someone/it V_1s then there is a state of N_1. d) if there is a state of N_1 then someone/it is Adj_1, and if someone/it is Adj_1 then there is a state of N_1. e) if someone/it V_1s then someone/it is Adj_1, and if someone/it is Adj_1, then someone/it V_1s.	{destruction}N **xpos near synonym** {destroy}V {destroy}V **xpos near synonym** {destruction}N {movement}N **xpos near synonym** {move}V {move}V **xpos near synonym** {movement}N {sleep}N **xpos near synonym** {sleep}V {sleep}V **xpos near synonym** {sleep}N {poverty}N **xpos near synonym** {poor}Adj {poor}Adj **xpos near synonym** {poverty}N {live}V **xpos near synonym** {alive}Adj {alive}Adj **xpos near synonym** {live}V
is XPOS near antonym (N, V and Adj)	a) if someone/it V_1s then N_1 does not take place, and if N_1 takes place then someone/it does not V_1. b) if someone/it V_1s then someone/it is not in a state of N_1, and if someone/it is in a state of N_1 then someone/it does not V_1s. c) if someone/it V_1s then someone/it is not Adj_1, and if someone/it is Adj_1 then someone/it does not V_1s.	{awakening}N **xpos near antonym** {fall asleep}V {fall asleep}V **xpos near antonym** {awakening}N {love}V **xpos near antonym** {hate}N {hate}N **xpos near antonym** {love}V {sleep}V **xpos near antonym** {awake}Adj {awake}Adj **xpos near antonym** {sleep}V

Relation	Tests	Examples
role/involved (N and V)	a) a N_1 is the one/that who/which is typically involved in V_1ing.	{vehicle}$_N$ **role** {travel}$_V$ {travel}$_V$ **involved** {vehicle}$_N$
	b) a N_1 is the one/that who/which does V_1, typically intentionally.	{teacher}$_N$ **role agent** {teach}$_V$ {teach}$_V$ **involved agent** {teacher}$_N$
	c) a N_1 is the one/that who/which undergoes V_1.	{student}$_N$ **role patient** {teach}$_V$ {teach}$_V$ **involved patient** {student}$_N$
	d) a N_1 comes into existence as a result of V_1, is the result of V_1, is created by V_1.	{crystal}$_N$ **role result** {crystallize}$_V$ {crystallize}$_V$ **involved result** {crystal}$_N$
	e) a N_1 is the instrument/what is used to V_1.	{pen}$_N$ **role instrument** {write}$_V$ {write}$_V$ **involved instrument** {pen}$_N$
	f) a N_1 is the place where V_1 happens.	{school}$_N$ **role location** {teach}$_V$ {teach}$_V$ **involved location** {school}$_N$
	g) a N_1 is the place from where V_1ing begins/starts/happens/ one V_1s.	{start}$_N$ **role source direction** {race}$_V$ {race}$_V$ **involved source direction** {start}$_N$
	h) a N_1 is the place to which V_1ing happens/ one V_1s.	{prison}$_N$ **role target direction** {incarcerate}$_V$ {incarcerate}$_V$ **involved target direction** {prison}$_N$
co_role (N)	a) if both N_1 and N_2 are in a role relation with V_1, then a co_role relations holds between N_1 and N_2.	{vehicle}$_N$ role {transport}$_V$ {goods}$_N$ role {transport}$_V$ {vehicle}$_N$ **co role** {goods}$_N$ {goods}$_N$ **co role** {vehicle}$_N$
	b) if N_1 is in a role agent relation with V_1, and N_2 is in a role patient relation with V_1, then a co_agent patient/co_patient agent relation holds between N_1 and N_2.	{teacher}$_N$ role agent {teach}$_V$ {student}$_N$ role patient {teach}$_V$ {teacher}$_N$ **co agent patient** {student}$_N$ {student}$_N$ **co patient agent** {teacher}$_N$
	c) if N_1 is in a role agent relation with V_1, and N_2 is in a role instrument relation with V_1, then a co_agent instrument/co_instrument agent relation holds between N_1 and N_2.	{photographer}$_N$ role agent {photograph}$_V$ {camera}$_N$ role instrument {photograph}$_V$ {photographer}$_N$ **co agent instrument** {camera}$_N$ {camera}$_N$ **co instrument agent** {photographer}$_N$
	d) if N_1 is in a role agent relation with V_1, and N_2 is in a role result relation with V_1, then a co_agent result/co_result agent relation holds between N_1 and N_2.	{photographer}$_N$ role agent {photograph}$_V$ {photo}$_N$ role result {photograph}$_V$ {photographer}$_N$ **co agent result** {photo}$_N$ {photo}$_N$ **co result agent** {photographer}$_N$
	e) if N_1 is in a role patient relation with V_1, and N_2 is in a role instrument relation with V_2, then a co_patient instrument/co_instrument patient relation holds between N_1 and N_2.	{ice}$_N$ role patient {cut}$_V$ {ice saw}$_N$ role instrument {cut}$_V$ {ice}$_N$ **co patient instrument** {ice saw}$_N$ {ice saw}$_N$ **co instrument patient** {ice}$_N$
	f) if N_1 is in a role patient relation with V_1, and N_2 is in a role result relation with V_1, then a co_patient result/co_result patient relation holds between N_1 and N_2.	{pastry dough}$_N$ role patient {bake}$_V$ {pastry}$_N$ role result {bake}$_V$ {pastry dough}$_N$ **co patient result**{pastry}$_N$ {pastry}$_N$ **co result patient** {pastry dough}$_N$
	g) if N_1 is in a role result relation with V_1, and N_2 is in a role instrument relation with V_1, then a co_result_instrument/co_instrument_result relation holds between N_1 and N_2.	{photo}$_N$ role result {photograph}$_V$ {camera}$_N$ role instrument {photograph}$_V$ {photo}$_N$ **co result instrument** {camera}$_N$ {camera}$_N$ **co instrument result** {photo}$_N$
in manner/ manner of (V and Adv)	to V_1 is to V_2 in an Adv_1 manner	{slurp}$_V$ **in manner** {noisily}$_{Adv}$ {noisily}$_{Adv}$ **manner of** {slurp}$_V$
be in state/ state of (N and Adj)	an N_1 is the one/that to whom/which the state Adj_1 applies	{infant}N **be in state** {young}$_{Adj}$ {young}$_{Adj}$ **state of** {infant}$_N$ *reversed*
belongs to class/ has instance (Proper N and N)	ProperN1 is one of the N_1s, but N_1 is not one of the ProperN1s.	{London}$_{ProperN}$ **belongs to class** {city}$_N$ {city}$_N$ **has instance** {London}$_{ProperN}$

Table 4: ENW relations available between synsets of the same or different POS

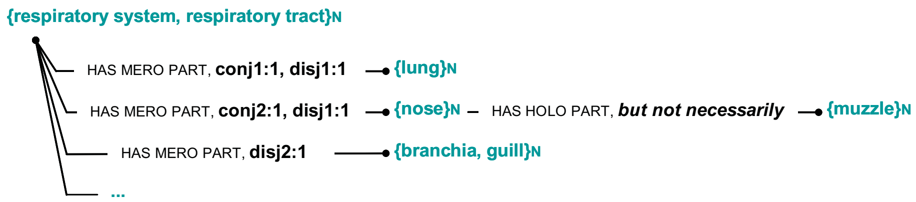
Table 4 contains, in fact, the relations considered in EWN that differentiate it from WN 1.5. On the one hand, these relations are, or can be, established between different POS, given the distinction in terms of 1st, 2nd and 3rd order entities, and reflecting the uniform treatment the model gives to all POS. On the other, these relations also reflect a greater degree of specification of the lexical-conceptual relations that link the nodes in the lexicon.

The SUBEVENT and CAUSE relations correspond to asubdivision of the WN 1.5 ENTAILMENT relation, allowing a richer characterization of the relations holding between events. ROLE and CO-ROLE relations constitute, respectively, the implementation of relations holding between events and their typical participants, and between participants themselves. IN MANNER and IN STATE relations allow the linking of adjectives and adverbs to verbs and nouns. Finally, the HAS INSTANCE relation links proper nouns to the concepts (synsets) they instantiate. This relation allows for including proper nouns in the general lexicon, which is quite interesting for inference purposes, named entity identification, among several other applications.

The relations presented in Table 4, as well as the MERONYM/HOLONYM relation in Table 3, are typically asymmetric relations. In order to further specify the relations established, EWN proposes a small set of relation labels, separately or jointly applied to EWN links: *conjunction*, *disjunction*, *negation*, *non-factive* and *reversed*.

Conjunction and *disjunction* labels are typically used to differentiate multiple relations of the same type associated to a given synset. These labels distinguish, for instance, mandatory – typically definitional – constituent parts of a given entity. *Reversed* label indicates by default the counterpart of relations, and can be read as *but not necessarily*.

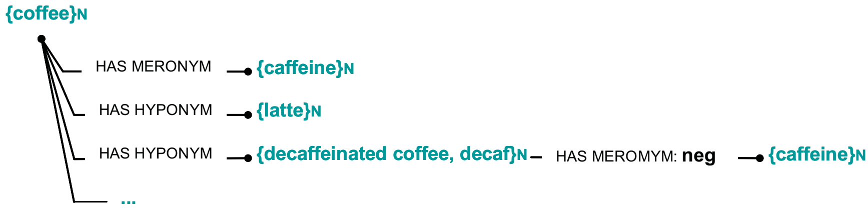
(14) Examples of *conjunction*, *disjunction* and *but not necessarily* labels



In (14) is stated that a *respiratory system* is composed of *lung* and *nose* conjointly, or, in alternative, of *branchia*, but *lung* and *nose* and *branchia* cannot be simultaneously part of *respiratory system*. On its turn, a *nose* can be part of a *muzzle*, but not necessarily.

Negation label is used to explicitly express that a given relation does not hold. This label is used, for instance, to tag a prototypical relation established at the hyperonym level that is not inherited by a given hyponym, as exemplified by the synsets in (15), where {decaffeinated coffee, decaf}$_N$ is a kind of {coffee}$_N$, but it is a coffee without {caffeine}$_N$ (example adapted from Marrafa (2001:50)).

(15) Example of *negation* label



The *non-factive* label is used to differentiate several types of lexical entailment expressed by the subevent and cause relations. In the wordnets, ENTAILMENT, CAUSE and SUBEVENT relations (ENTAILMENT in WN 1.5 and CAUSE and SUBEVENT in EWN) express the relations holding between two events that are temporally disjoint or that partially overlap in time, since when there is entailment and temporal co-extensiveness the relation linking two concepts it is typically HYPERONYMY/HYPONYMY. As stated above, *reversed* label (*but not necessarily*) is ascribed by default to the converse direction of an asymmetric relation encoded in the database. *Non-factive* label, on its turn, marks the case where a given causal relation between two events does not necessarily hold, being otherwise assumed that the relation is factive.

(16) Example of *non-factive* labels (taken from Vossen (2002:34,36))

a. {murder}$_V$ CAUSES {die}$_V$ (factive)
{die}$_V$ IS CAUSED BY {murder}$_V$ ***but not necessarily***
(to *murder* causes to *die*, but to *die* is not necessarily caused by to *murder*)

b. {search}$_V$ CAUSES {find}$_V$ **non-factive**
{find}$_V$ IS CAUSED BY {search}$_V$ **non-factive**
(to *search* may cause to *find*, and to *find* may be caused by to *search*)

c. {snore} IS SUBEVENT OF {sleep}
{sleep} HAS SUBEVENT {snore} ***but not necessarily***
(*to snore* takes place when someone *sleeps*, but whenever someone *sleeps* *snoring* does not necessarily takes place.)

Finally, and in what concerns variant description, EWN also provides a quite large set of usage features regarding variant register, style, etc., each with several possible values (formal, informal, technical, and so on); some morphosyntactic features (tense, gender, number, etc.), definitions and examples, which potentiate the use of the database by non-specialist users, such as language students, translators and so on.

As mentioned earlier in this subsection, the multilingual character of EWN does not impose strict constraints on the development of the databases for particular languages. The EWN architecture assures a high level of independence to each language, both in terms of enlargement strategies and of modifications and enhancement of lexical semantic specifications. Moreover, the EWN model itself – as the wordnet model, in general – is continuously absorbing and profiting from state of the art research, typically developed within the projects for particular languages. Over the years, wordnet databases have been consistently used to improve semantic processing tasks in language-based applications, reflecting also concerns of wordnet developers on the improvement of the usability of the model (see Marrafa & Mendes (2007), Clark *et al.* (2008)).

The following section will, thus, be dedicated to the description of the Portuguese WordNet (henceforth WN.PT) in what concerns its distinguishing features.

2.1.3 WordNet.PT

WN.PT is the lexical conceptual database for Portuguese (Marrafa 2001, Marrafa 2002)[2]. Developed according to the general approach of EWN, WN.PT presents some distinctive properties concerning, on the one hand, its building options and lexical coverage strategies and, on the other, the extension of the set of relations used.

The data in WN.PT are all manually selected, described and encoded, resulting in a smaller lexical resource, compared with automatically and semi-automatically constructed databases, but a more reliable one and with denser nets of relations. The enlargement of the database has mainly followed the semantic domains approach, involvind the encoding of lexical items (regardless of their POS) from particular semantic areas, such as food, clothing, living beings, transportation, health, sciences, sports, and so on. The semantic domain division, however, does not result in independent and hermetic chunks of the lexicon since the methodology followed intends a wide coverage of the common lexicon of Portuguese and all the synsets defined in the database are related to all the pertinent nodes, regardless of the semantic domain. For instance *bacterium* is encoded as a type of *living being*, but it is also related to *antibiotic drug* and *infection*, which clearly fall under the health/medicine semantic field.

This development strategy has provided the opportunity of testing and implementing new cross POS relations, mainly in what concerns the relational implementation of adjectives, but also in what concerns the event structures of verbs and nouns, and continues to provide grounds for further research on Lexical Semantics (Mendes & Chaves (2001), Marrafa *et al.* (2006), Amaro *et al.* (2006), Marrafa & Mendes (2006), for instance).

[2] WN.PT has been developed since 1998 at the Center of Linguistics of the University of Lisbon by CLG – Group for the Computation of Lexical and Grammatical Knowledge, coordinated by Professor Palmira Marrafa, within the context of several research projects (http://www.clul.ul.pt/clg/wordnetpt/index.html).

WN.PT adopts almost entirely the set of relations defined in the EWN model. Exceptions are made to the relations DERIVES/IS DERIVED, PERTAINS/IS PERTAINED and BE IN STATE/STATE OF. The first two relations mentioned are morphological relations used to link variants, representing no semantic information. In fact, these relations are only to be used if there is some other semantic relation linking the two variants or the synsets in which they are included (Vossen 2002: 37), hence being somewhat complementary. The BE IN STATE/STATE OF relation links nouns referring entities that are in a particular state expressed by a given adjective, as for instance {rich}$_{N}$ and {rich}$_{Adj}$.

Interesting enough, the three mentioned relations are used to co-relate nouns and adjectives, and correspond to the need to link adjectives in the lexicon, since the structuring relation of hyperonymy/hyponymy is not very productive for this POS.

Fundamental research on adjectives and on event structure developed within the WN.PT project (Marrafa 2005, Mendes 2006) has led to the determination of three new semantic relations that further support discarding of the previously mentioned morphological relations. In WN.PT adjectives are linked to nouns and verbs through:

(i) the **CHARACTERIZES WITH REGARD TO/CAN BE CHARACTERIZED** relation (linking descriptive adjectives to the attribute they modify);

(17) Examples:

{tall}$_{Adj}$ CHARACTERIZES WITH REGARD TO {height}$_{N}$
{short}$_{Adj}$ CHARACTERIZES WITH REGARD TO {height}$_{N}$

{height}$_{N}$ CAN BE CHARACTERIZED BY {tall}$_{Adj}$
CAN BE CHARACTERIZED BY {short}$_{Adj}$

{poor}$_{Adj}$ CHARACTERIZES WITH REGARD TO {poverty}$_{N}$
{poverty}$_{N}$ CAN BE CHARACTERIZED BY {poor}$_{Adj}$

(ii) the **IS RELATED TO** relation (allowing for the specification of the set of properties introduced by relational adjectives by linking them to the in the net lexicalizing this set of properties);

(18) Examples:

{marine}$_{Adj}$ IS RELATED TO {sea}$_{N}$
{sea}$_{N}$ IS RELATED TO {marine}$_{Adj}$

{atomic}$_{Adj}$ IS RELATED TO {atom}$_{N}$
{atom}$_{N}$ IS RELATED TO {atomic}$_{Adj}$

(iii) the **IS CHARACTERISTIC OF/HAS AS A CHARACTERISTIC TO BE** relation (replacing the BE IN STATE/STATE OF relation, and allowing the encoding of salient characteristics of entities, providing grounds for the identification of typical semantic domains of adjective application);

(19) Examples:

{young}$_{Adj}$ IS CHARACTERISTIC OF {child}$_N$

{child}$_N$ HAS AS CHARACTERISTIC TO BE {young}$_{Adj}$

{marine}$_{Adj}$ IS CHARACTERISTIC OF {seal}$_N$

{seal}$_N$ HAS AS CHARACTERISTIC TO BE {marine}$_{Adj}$

(iv) the **HAS TELIC SUBEVENT/IS TELIC SUBEVENT** relation, on its turn, allows for the encoding of telicity of LCS-deficitary verbs (see Marrafa (2005).

(20) Examples:

{sadden}$_V$ HAS TELIC SUBEVENT {sad}$_{Adj}$

{sad}$_{Adj}$ IS TELIC SUBEVENT {sadden}$_V$ *but not necessarily*

This new set of relations replaces derivational and morphological based relations, reflecting semantic properties of adjectives, nouns and verbs not considered in the EWN model.

Relations	**WN 1.5**	**EWN**	**WN.PT**
similar to	✓	✗	✗
participle	✓	✗	✗
synonym	✓	✓	✓
hyponym/hyperonym	✓	✓	✓
antonym	✓	✓	✓
meronym/holonym	✓	✓	✓
causes/is caused by	✓	✓	✓
is subevent/has subevent	✓ *entailment*	✓	✓
XPOS hyponym/XPOS hyperonym	✗	✓	✗
derives/is derived	✓	✓	✗
pertains/ is pertained	✓ *relational adjective*	✓	✗
near synonym	✗	✓	✓
XPOS near synonym	✗	✓	✓
near antonym	✗	✓	✓
role/involved	✗	✓	✓
co role	✗	✓	✓
in manner/manner of	✗	✓	✓
instance of/has instance	✗	✓	✓
be in state/state of	✓ *is value of/ attribute*	✓	✓ is characteristic of/has as a characteristic to be
characterizes with regard to/ can be characterized by	✗	✗	✓
is related to	✗	✗	✓
has Telic subevent/is Telic subevent	✗	✗	✓
Total of relations	**11**	**17**	**17**

Table 5: Relations available in WN 1.5, EWN and WN.PT

Table 5, above, presents the relations used in the three databases, WN 1.5, EWN and WN.PT, makking apparent their major differences, but also, and more importantly, the continuity of the model.

In what concerns the verbal lexicon, and given that the majority of the relations defined considers no constraints on POS (with the natural exceptions of the distinction between part-whole relations for nouns (MERONYMY/HOLONYMY), and the IS CHARACTERISTIC OF/HAS AS CHARACTERISTIC TO BE relation that links adjectives and nouns, and which can be regarded as the correspondent relation of IN MANNER/MANNER OF which links adverbs and verbs), verbs can be encoded using virtually any of the relations presented above. However, and as stated before, the first tasks required for the construction of a relational net concern the organization of the lexical items in terms of the two structuring relations of SYNONYMY and HYPERONYMY/HYPONYMY. A brief report on the construction of the relational net for Portuguese verbs of movement is presented in the next section.

2.2 Verbs of movement in WordNet.PT

The construction of a relational net for Portuguese verbs of movement followed, necessarily, the semantic domain approach: starting from common verbs of movement as *mover* (move, change location), *mexer* (move, change position), *ir* (go), *correr* (run), *balançar* (sway), etc., a first and open list of verbs was compiled. From the analysis and treatment of this list, several other verbs were added, by natural lexical association: *andar* (walk) – *coxear* (limp), *rastejar* (crawl), *nadar* (swim), *voar* (fly); *ir* (go) – *vir* (come), *voltar* (come back), *regressar* (return); *pôr* (put) – *tirar* (take); *abanar* (shake) – *oscilar* (oscillate), *tremer* (tremble); and so on.

The organization of the verbs listed in synonym sets and, more importantly, the establishment of hyperonymy/hyponymy relations between the synsets, required the definition of the higher nodes of the network, in order to assure conceptual consistency throughout the net and to test the inheritance of the concepts denoted from the higher to the lower nodes in the hierarchy. This process revealed several issues concerning the semantic and syntactic diversity of directly related verbs within the net, leading to the need to establish and motivate some encoding options.

2.2.1 Top nodes

As presented in the previous chapter, movement verbs denote events in which there is change of location, either change of location of a given entity as a whole (change of location verbs), or change of location of some part of an entity (change of position verbs). The core concept of movement, as used in physics, for instance, is defined as the "change with time of the position or orientation of a body"[3]. This notion subsumes a more abstract notion of movement that, if lexicalized, could constitute the initial node for the relational net for verbs of movement. In Portuguese, sentences such as the ones in (22) seem to indicate that the verb *mexer* could be this top node:

(21) Examples taken from CRPC – Reference Corpus of Contemporary Portuguese[4]

a. *A questão não era discutir se estava realmente morto ou ainda* ***mexia****, a questão era saber quem o queria matar e o que fazer para o salvar*.

Corpus RLD, Ref: J60763

(The issue was not to discuss whether he was really dead or if he **moved** still, the issue was knowing who wanted to kill him and what to do to save him.) ≅ he did not stand/remain still

b. *… e atirar sobre tudo o que* ***mexe****, mesmo que a época só abra para as rolas, os patos e os pombos.* Corpus RLD, Ref: J42377

(… and shoot everything that **moves** even if the (hunting) season only opens for turtledoves, ducks and pigeons.) ≅ everything that does not stand/remain still

The verb *mexer* in these examples seems to denote 'change the position or orientation' of a given entity, or in other words, change the state of rest of a given entity. The follow through of this hypothesis should demonstrate that *mexer* (≅ change the state of rest) can be the hyperonym of the synset $\{\text{mover, deslocar}\}_V$ (*move* ≅ change location). However, the results of the tests are not straightforward[5]:

(22) $\{\text{deslocar}\}_V$ is hyponym of $\{\text{mexer}\}_V$, if

a. #*deslocar* é *mexer* alterando a localização, mas *mexer* não é *deslocar* sem alterar a localização

(*move* (≅ change the location) is *move* (≅ change the state of rest) by changing the location but *move* (≅ change the state of rest) is not *move* (≅ change the location) without changing the location)

b. *deslocar* entails *mexer* but *mexer* does not entail *deslocar*

#quando alguém *desloca* algo então alguém *mexe* algo mas quando alguém *mexe* algo, alguém não *desloca* algo obrigatoriamente

---

[3] in *Encyclopedia Britannica* (http://www.britannica.com/EBchecked/topic/394061/motion).

[4] CRPC – Reference Corpus of Contemporary Portuguese is hosted at the Center of Linguistics of the University of Lisbon and is available for online queries in http://www.clul.ul.pt/english/sectores/linguistica_de_corpus/projecto_rld_pesquisa_PE.php.

[5] For transparency reasons, explored further ahead in this section, we will use only the variant 'deslocar' in the tests presented in (22) and (23).

(when someone *moves* (≅ changes the location of) something then someone *moves* (≅ changes the state of rest of) something; but when someone *moves* (≅ changes the state of rest of) something then someone does not necessarily *moves* (≅ changes the location of) something)

c. *deslocar* is not temporally coextensive with *mexer*.
#alguém está necessariamente a *mexer* alguma coisa durante todo o tempo em que alguém está a *deslocar* alguma coisa (mas alguém não está necessariamente a *deslocar* alguma coisa durante todo o tempo em que alguém está a *mexer* alguma coisa).
(someone is necessarily *moving* (≅ changing the state of rest of) something for as long as someone is *moving* (≅ changing the location of) something (but someone is not necessarily *moving* (≅ changing the location of) something for as long as someone is *moving* (≅ changing the state of rest of) something).

Also, the contexts used for testing the hierarchical nature of the hyperonymy relation do not present the expected results:

(23) a. #Ele *deslocou* a caixa, *mexendo*-a assim para a ver melhor.
(He moved (≅ changed the location of) the box, moving (≅ changing the state of rest of) it so to get a better look at it.)

b. Ele *deslocou* a caixa mas não a *mexeu*.
(He moved (≅ changed the location of) but did not moved (≅ changed the state of rest of) it.)

Given the subspecification level of the concepts denoted by the verbs in (24) and (25) and the evident correlation between them, it is not strange that the verbs *mexer* and *mover*, for instance, are almost interchangeable in sentences such as the ones presented in (26).

(24) <u>Change of location</u>:
For a given entity E; for the locations $L_1, ..., L_n$; at a given time span $\{t_1, ..., t_n\}$;
E changes location iff:
E in t_1 is in L_1,
E in t_n is in L_n, and
$L_1 \neq L_n$.

(25) Change of position:

For a given entity E; for the locations $L'_1, ..., L'_n$ in a given location L; at a given time span $\{t_1, ..., t_n\}$;

E changes position iff:

$L'_1, ..., L'_n \subset L$;

E (or part of E) in t_1 is in L'_1,

E (or part of E) in t_n is in L'_n,

$L'_1 \neq L'_n$.

(26) a. A criança **mexia** as pernas, os braços e cabeça. *and* A criança **movia** as pernas, os braços, a cabeça.
(The child moved the legs, arms and head.)

b. Tentou sorrir-lhe, mas apenas **mexeu** os lábios. *and* Tentou sorrir-lhe, mas apenas **moveu** os lábios.
(He tried to smile, but he only moved the lips.)

Note, however, that this interchangeability does not hold with the verb *deslocar*, synonym of *mover* (move ≅ change location), raising the issue of whether or not these two verbs – *mover* and *deslocar* – are in fact synonyms.

(27) a. #A criança **deslocava** as pernas, os braços e a cabeça.
(The child displaced the legs, arms and head)

b. #Tentou sorrir, mas apenas **deslocou** os lábios.
(He tried to smile, but he only displaced the lips)

Interesting enough, the only contexts in which *mover* and *deslocar* are not replaceable by each other are the ones involving body parts. *Deslocar* typically occurs with articulated body parts, meaning to become out of joint/dislocate (changing, in fact, the proper location):

(28) a. A criança **deslocou** o braço/o ombro/a anca.
(The child dislocated the arm/shoulder/hip)

b. #A criança **deslocou** a cabeça/os lábios/os olhos.
(The child dislocated the head/lips/eyes)

Mover, on its turn, can occur with all the naturally moving body parts, typically articulated but not necessarily, but does not convey the concept of becoming out of joint/dislocate. It does not, however, occur with less moveable parts, such as *nose*.

(29) a. A criança **moveu** o braço/?o ombro/a anca.
(The child moved the arm/shoulder/hip)

b. A criança **moveu** a cabeça/os lábios/os olhos/?o nariz.
(The child moved the head/lips/eyes/nose)

These examples seem to corroborate the hypothesis of having three distinct synonym sets, conveying different concepts, the use of the variants involved also being conditioned by ambiguity issues, resulting thus in preferred readings when in specific contexts.

(30) a. $\{$mexer$\}_V$ [move ≅ change position]

b. $\{$mover, deslocar$\}_V$ [move ≅ change location]

c. $\{$deslocar$\}_V$ [dislocate ≅ put, bones, out of place]

Mexer and *mover* can occur in similar contexts when referring to naturally moveable body parts, since a change of position event refers to a specific change of location event, being, thus, the case of conveying a change of location within a reference frame, e.g. *O rapaz mexeu o braço, mas manteve-o à altura dos ombros*. (The boy moved the arm, but kept it at shoulder height), or the case of conveying the change of location of a moveable part, e.g. *O rapaz moveu o braço para baixo, não o conseguindo manter à altura dos ombros*. (The boy moved his arm down, failing to keep it at shoulder height). The replacement is not possible with non moveable (in the sense of changing location) body parts such as *nose*: *O rapaz mexeu/?moveu o nariz por causa da comichão*. (The boy moved (≅ change position)/moved (≅ change location) his nose because it was itchy).

In the case of moveable body parts, typically the variant *mover* is preferred to the variant *deslocar* since the second verb can also denote put bones out of place, resulting in odd sentences. However, in larger and less ambiguous contexts, the replacement is possible:

(31) A criança moveu/deslocou a cabeça ligeiramente para a esquerda, de modo a ver melhor o palco. (cf. (28)b)
(The child moved/displaced his head slightly to the left to better see the stage)

Finally, *mexer* does not describe a situation necessarily involving the displacement of an entity/object. The 'change of location' movement requires the initial location in which the entity starts the movement to be different from the final location in which the movement ends: the entity moves from location A to location B (B ≠ A). The 'change of position' movement involves displacement of some part of the entity (that can constitute the whole entity) with respect to locations contained within the frame location L: some part of the entity (or all of it) moves within L. We cannot infer from a sentence like (32) that there was a change of location of the whole object:

(32) O vento mexe as árvores.
(The wind moves the trees)

In sum, we consider the synsets {mover, deslocar}$_V$ [≅ change location; move, displace] and {mexer}$_V$ [≅ change position; move] as the top nodes of the wordnet of Portuguese verbs of movement since there is no lexicalization of a more underspecified concept (change the state of rest, for instance) that could refer to both types of movement events.

2.2.2 Word-sense differentiation

In WN 1.5, the subnets for change of location verbs and for change of position verbs are also subdivided in two independent structures, distinguishing the causative senses of the verbs, (33)a, from the inchoative senses, (33)b, (see Fellbaum (1998: 76)). The two senses are linked to each other by a CAUSE relation, (33)c.

(33) a. {move 1}$_V$ [cause to move]

b. {move 2}$_V$ [move/change location]

c. {move 1}$_V$ CAUSES {move 2}$_V$
{move 2}$_V$ IS CAUSED BY {move 1}$_V$

The distinction between these two senses and the option to encode them in two different entries seems also to be related to the number of arguments and the syntactic realization of each sense. {move 1}$_V$ refers to a two-place predicate whose first argument causes the second argument to endure the movement event, in (34)a, whereas {move 2}$_V$ refers to a one-place predicate whose only argument endures the movement event, see (34)b.

(34) a. John moved the chair.

b. The chair moved.
b.' John moved.

This strategy of multiplying the number of entries according to the number possible senses a given word can have in context fails to grasp the differences between the sentences in (34)b, an inchoative structure, and (34)b' an unergative structure. CAUSE relation in WN 1.5 is used to link unrelated pairs of verbs or synsets, such as *show-see*, *fell-fall*, *expel-leave*, but also verbs that enter causative/inchoative alternations like *break-break* (see Fellbaum (1998a: 83-84)), accounting for issues of different order.

In the set of Portuguese verbs of movement we have also to consider this issue. As demonstrated in (34)a, one-place verbs of movement do not necessarily correspond to the inchoative construction of causative verbs. The examples in (35) corroborate this statement.

(35) a. *A cadeira arrastou-se. → Algo/Alguém arrastou a cadeira. : *false*
(The chair dragged itself→ Something/someone dragged the chair.)

a′. A criança arrastou-se. → Algo/Alguém arrastou a criança. : *false*
(The child dragged himself → Something/Someone dragged the child)

b. A cadeira recuou ?#(com o vento). → O vento recuou a cadeira. : *true*
(The chair moved back (with the wind) → The wind moved back the chair)

b′. A criança recuou. → Algo/Alguém recuou a criança. : *false*
(The child moved back → Something/Someone moved back the child)

c. O barril rebolou/rolou. → Algo/Alguém rebolou/rolou o barril. : *true*
(The barrel rolled.→ Something/Someone rolled the barrel)

c′. A criança rebolou/rebolou-se/rolou. → Algo/Alguém rebolou/rolou a criança. : *false*
(The child rolled → Something/Someone rolled the child)

The application of the CAUSE relation tests to the verbs considered above is also not conclusive:

(36) a. {mover, deslocar}$_V$ causes {mover-se/deslocar-se}$_V$ iff:

i) the event of *mover/deslocar* causes the event of *mover-se/deslocar-se* to take place, OR the event of *mover/deslocar* has the event of *mover-se/deslocar-se* as a consequence, OR the event of *mover/deslocar* leads to the event of *mover-se/deslocar-se*, but not the converse: **not necessarily**

ii) the event of *mover/deslocar* may cause the event of *mover-se/deslocar-se*, OR the event of *mover/deslocar* may have the event of *mover-se/deslocar-se* as a consequence, OR the event of *mover/deslocar* may lead to the event of *mover-se/deslocar-se*, but not the converse.: **true**

iii) *mover/deslocar* is not hyperonym of *mover-se/deslocar-se*, and
if the event of *mover/deslocar* takes place it causes/may cause the event of *mover-se/deslocar-se* to take place at the same time, but not the converse.: **not necessarily**

b. {arrastar}$_V$ causes {arrastar-se}$_V$ iff:

i) the event of *arrastar* causes the event of *arrastar-se* to take place, OR the event of *arrastar* has the event of *arrastar-se* as a consequence, OR the event of *arrastar* leads to the event of *arrastar-se*, but not the converse: **false**

ii) the event of *arrastar* may cause the event of *arrastar-se*, OR the event of *arrastar* may have the event of *arrastar-se* as a consequence, OR the event of *arrastar* may lead to the event of *arrastar-se*, but not the converse.: **false**

iii) *arrastar* is not hyperonym of *arrastar-se*, and
if the event of *arrastar* takes place it causes/may cause the event of *arrastar-se* to take place at the same time, but not the converse.: **false**

c. {recuar 1}$_V$ causes {recuar 2}$_V$ iff:

i) the event of *recuar 1* causes the event of *recuar 2* to take place, OR the event of *recuar 1* has the event of *recuar 2* as a consequence, OR the event of *recuar 1* leads to the event of *recuar 2*, but not the converse: **not necessarily**

ii) the event of *recuar 1* may cause the event of *recuar 2*, OR the event of *recuar 1* may have the event of *recuar 2* as a consequence, OR the event of *recuar 1* may lead to the event of *recuar 2*, but not the converse.: **true**

iii) *recuar 1* is not hyperonym of *recuar 2*, and
if the event of *recuar 1* takes place it causes/may cause the event of *recuar 2* to take place at the same time, but not the converse: **not necessarily**

d. {rebolar, rolar}$_V$ causes {rebolar, rebolar-se, rolar}$_V$ iff:

i) the event of *rebolar/rolar* causes the event of *rebolar/rebolar-se/rolar* to take place OR the event of *rebolar/rolar* has the event of *rebolar/rebolar-se/rolar* as a consequence OR the event of *rebolar/rolar* leads to the event of *rebolar/rebolar-se/rolar*, but not the converse: **not necessarily**

ii) the event of *rebolar/rolar* may cause the event of *rebolar/rebolar-se/rolar*, OR the event of *rebolar/rolar* may have the event of *rebolar/rebolar-se/rolar* as a consequence, OR the event of *rebolar/rolar* may lead to the event of *rebolar/rebolar-se/rolar*, but not the converse.: **false**

iii) *rebolar/rolar* is not hyperonym of *rebolar/rebolar-se/rolar*, and
if the event of *rebolar/rolar* takes place it causes/may cause the event of *rebolar/rebolar-se/rolar* to take place at the same time, but not the converse: **not necessarily**

In what concerns the occurrence in causative/inchoative alternations, and as demonstrated above, the verbs presented above show significant differences. Some verbs do not allow inchoative constructions (*arrastar* (≅ drag)); some allow inchoative constructions but seem to require the explicit statement of the cause of the event (*recuar* (≅ move back)); and some license either the explicit statement or the omission of the cause of the event (*mover/deslocar* (≅ cause change of location) and *rebolar/rolar* (≅ roll)).

In fact, there are some issues to ponder before considering these verbs as causative/inchoative pairs. According to Haspelmath (1993:90) "An inchoative/causative verb pair is defined semantically: it is a pair of verbs which express the same basic situation (...) and differ only in that the causative verb meaning includes an agent participant who causes the situation, whereas the inchoative verb meaning excludes a causing agent and presents the situation as occurring spontaneously." At the light of this reasoning, the event denoted by the verb *arrastar* (≅ drag) cannot occur spontaneously, explaining thus its inability to occur in causative/inchoative alternations. However, the same seems to be true for the verb *recuar* (≅ move back), since the acceptability of the inchoative construction seems dependent on the realization of the PP expressing the cause of the event. In what concerns the question in hand – the differentiation of verb senses and respective encoding in a relational model – it also seems odd that the alternation is not extensive to the entire verbal paradigm, weakening the motivation for the division of the net in two separate structures.

The use of separate lexical entries to distinguish the causative and the non causative senses of a given verb is a strategy not adopted in WN.PT. As argued in Marrafa & Mendes (2007), verbs that enter this alternation always denote the same event both in causative sentences and in inchoative ones. The inchoative sentence does not necessarily refer to the final state of the transition event denoted by the verb, as the co-occurrence with progressive temporal adverbs, such as *lentamente* (≅ slowly), demonstrates (see (37), examples taken from Marrafa & Mendes 2007), and thus the motivation for the differentiation of senses is overcome:

(37) a. *Ele aqueceu a sopa lentamente*. $_{[\text{CAUSATIVE}]}$
(He warmed the soup slowly)

b. *A sopa aqueceu lentamente*. $_{[\text{INCHOATIVE}]}$
(The soup warmed slowly)

Both (37)a and (37)b denote a process which precedes and implies a change of state of the same argument, *soup*, implying a state of affairs where the soup is warm.

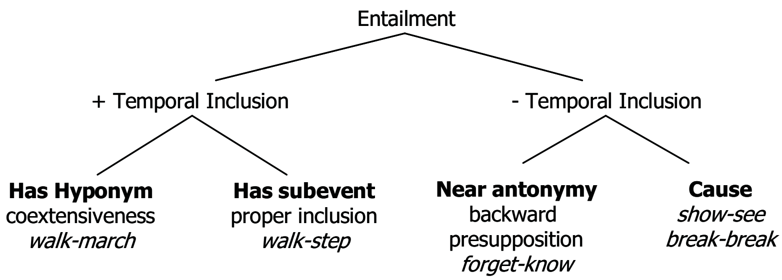
Given that the verbs in study do not entirely satisfy the tests for the establishment of a CAUSE relation between causative and non causative senses, and that the occurrence in causative/inchoative alternations does not require distinct entries, we will not consider distinct subnets for Portuguese verbs of movement.

Moreover, the definition of the causative sense for verbs of movement as the concept *cause to move* seems more accurate to define the concept expressed by synsets such as {atirar}$_V$ (≅ throw) or {disparar}$_V$ (≅ shoot), typically not acknowledged to be verbs of movement. We assume that the movement concepts of *change of location* and *change of position* refer to change of state events which can involve an external or internal cause, reflected in the argument structure of verbs.

2.2.3 Encoding options

Considering the typology of entailment relations expressed in wordnet models in (38), below, it seems that the entailment relation established between these verbs can also be consistent with a HYPERONYMY/HYPONYMY relation, given the conceptual proximity and the temporal inclusion established between the events denoted.

(38) Lexical entailment in WordNet



Adapted from Fellbaum (1998a:84)

It is thus possible to consider the synset {mover, deslocar}$_V$ (≅ change location) to be the hyperonym of {mover-se, deslocar-se}$_V$ (≅ change oneself/itself location):

(39) a. {mover, deslocar}$_V$ (≅ change location).

b. *mover-se*, *deslocar-se* é *mover, deslocar* a si próprio (to move is to change the location of oneself/itself).

Similarly to the relation of auto-HYPONYMY – in which the incorporation of an argument results in a different meaning of a given verb, related by hyponymy to the base concept denoting verb, e.g. {drink 1}$_V$ [ingest liquids] HAS HYPONYM {drink 2}$_V$ [drink alcoholic beverages] (see Fellbaum (1998: 86)) –, the hyponymy relation between the synsets {mover, deslocar}$_V$ (≅ change location) and {mover-se, deslocar-se}$_V$ (≅ change location of oneself/itself) can result from the same process. {mover-se, deslocar-se}$_V$ (≅ change location of oneself/itself) further determines that the second argument of the verb must be co-referent with the first, in an incomplete incorporation process, as seems to be indicated by presence of the reflexive clitic –*se*.

In this approach, the lexical-conceptual net of verbs of movement contains only the two top nodes defined, {mover, deslocar}$_V$ (≅ change location) and {mexer}$_V$ (≅ change position), which respectively have as the synsets {mover-se, deslocar-se}$_V$ (≅ change location of oneself/itself) and {mexer-se}$_V$ (≅ change position of oneself/itself), as hyponyms, among others (see (40)).

(40) Hyponym subtrees for the Portuguese wordnet of verbs of movement

{mover, deslocar}v
(*change location*)

- ...
- **{arrastar}v** (**mover** with effort over a surface that offers significant opposition to movement; ≅ *drag*)
- **{recuar}v** (**mover** back; ≅ *move back, recede*)
- **{rebolar, rolar}v** (**mover**, by consecutive tumbles near a surface; ≅ *roll*)
- **{afastar}v** (**mover** aside, away; ≅ *move aside*)
- **{mover-se}v** (**mover** *oneself/itself*)
  - **{arrastar-se}v** (**mover-se** with effort over a surface that offers significant opposition to movement; ≅ *drag oneself/itself*)
  - **{recuar}v** (**mover-se** back; ≅ *move back*)
  - **{rebolar-se, rebolar, rolar}v** (**mover-se** by consecutive tumbles near a surface; ≅ *roll oneself/itself*)
  - **{andar, caminhar}v** (**mover-se** in the ground using limbs; ≅ *walk*)
  - **{voar}v** (**mover-se** in the air; ≅ *fly*)
  - **{ir, deslocar-se}v** (**mover-se** in a specific direction; ≅ *go*)
  - ...

{mexer}v
(*change position*)

- **{abanar}v** (**mexer** from one position to the opposite one with impulses in different directions; ≅ *shake*)
  - **{balançar, balancear}v** (**abanar** smoothly; ≅ *sway*)
    - **{embalar}v** (**balançar**, animated entities, to get them to sleep; ≅ *rock, cradle*)
- **{virar}v** (**mexer** to put in a different (typically opposite) orientation; ≅ *turn*)
- **{mexer-se}v** (**mexer** *oneself/itself*)
  - **{tremer}v** (**mexer-se** with subtle, quick and repeated convulsions; ≅ *tremble, shiver*)
  - **{cabecear}v** (**mexer-se**, letting the head fall due to sleepiness; ≅ *nod*)
  - **{contorcer-se}v** (**mexer-se**, bending the whole body or parts of it, typically in pain; ≅ *contort, squirm*)
  - **{mover-se}v** (**mexer-se**, the whole body or parts of it, voluntarily)
    - **{levantar-se, erguer-se}v** (**mover-se** up to a vertical position; ≅ *rise, stand up*)
    - **{espreguiçar-se}v** (**mover-se** stretching the muscles of the body to relax; ≅ *stretch, stretch out*)
    - **{esbracejar}v** (**mover-se** agitating the arms energetically; ≅ *move the arms, gesticulate*)
    - ...

This encoding strategy does not reflect the causative/non-causative alternations in which some of these verbs occur, which, as stated before, are directly related to the internal structure of the events. Two-place predicates can generally enter causative/inchoative alternations, whereas the verbs in the subnet of hyponyms of {mover-se, deslocar-se}$_V$ and {mexer-se}$_V$, which are one-place predicates, cannot. The later denote an internally caused event and cannot generally co-occur with cause denoting PPs, or in contexts which force the inchoative reading (see (41)). On the contrary, the two-place predicates hyponyms of the synsets {mover, deslocar}$_V$ and {mexer}$_V$, when in inchoative constructions, cannot co-occur with phrases conveying intention:

(41) a. O soldado caiu do tanque já inconsciente e rebolou/rolou pela encosta (com/por causa do peso).

(The soldier fell from the tank already unconscious and rolled down the hill (with/because of the weight))

b. *O soldado caiu do tanque já inconsciente e rebolou-se/rolou-se pela encosta (com/por causa do peso).

(The soldier fell from the tank already unconscious and rolled-SE down the hill (with/because of the weight))

(42) a. O soldado rebolou-se/rebolou/rolou pela estrada para evitar ser visto.

(The soldier rolled-SE/rolled down the street to avoid being seen.)

b. *O soldado rebolou-se/rolou-se com/por causa do peso pela estrada para evitar ser visto.

(The soldier rolled-SE/rolled down the street with/because of the weight to avoid being seen.)

The specification of the concept denoted by the hyperonym, which in these cases results in the incorporation of the second argument of the verb, also reflects some specification of the type of entities selected by verbs, since they have to have intrinsic characteristics that allow them to endure a movement event internally caused or to move voluntarily. The aspects that distinguish the concepts in the relational net for Portuguese verbs of movement will be explored in detail in the next chapter.

Still concerning encoding options, it is necessary to motivate the positioning of the verbs that exhibit argument incorporation in the relational net, since they can be hyponyms of synsets composed of corresponding two-place verbs, or hyponyms of synsets denoting more general concepts, as {mover-se, deslocar-se}$_V$ (≅ *change location of oneself/itself*) or {mexer-se}$_V$ (≅ *change position of oneself/itself*). Let us consider, for instance, the synsets {arrastar}$_V$ (≅ *drag*) and {arrastar-se}$_V$ (≅ *drag oneself/itself*). The hyperonymy tests seem to validate both cases:

(43) Hyperonymy test:

V_2 is hyponym of V_1 if:

a. V_2 is V_1+$AdvP_i/AdjP_j/NP_k/PP_l$, but V_1 is not V_2+ $AdvP_i/AdjP_j/NP_k/PP_l$; and

b. V_2 entails V_1.

(44) a. *arrastar-se* é *arrastar* a si próprio, mas *arrastar* não é *arrastar-se* a si próprio

(drag oneself/itself is *drag* oneself/itself, but *drag* is not to *drag* oneself/itself)

b. *arrastar-se* implica *arrastar* e sempre que alguém *se arrasta* esse alguém *arrasta*.

(*drag oneself* entails *drag* and someone/something is necessarily *dragging* for as long as someone/something is *dragging oneself*)

(45) a. *arrastar-se* é *mover-se* com esforço sobre uma superfície que oferece resistência significativa ao movimento, mas *mover-se* não é *arrastar-se* com esforço sobre uma superfície que oferece resistência significativa ao movimento

(*drag oneself/itself* is *move oneself/itself* with effort over a surface that offers significant opposition to movement, but *move oneself/itself* is not to *drag oneself/itself* with effort over a surface that offers significant opposition to movement

b. *arrastar-se* implica *mover-se* e sempre que alguém *se arrasta* esse alguém *move-se*.

(*drag oneself/itself* entails *move oneself/itself* and someone/something is necessarily *moving oneself/itself* for as long as someone/something is *dragging oneself/itself*)

Although both synsets pass the hyperonymy tests, the resulting sentences for the synset {mover-se}$_V$ (≅ move oneself/itself) are more easily accepted. Being particularly redundant the resulting sentences for the synset {arrastar}$_V$ (≅ drag) are not so easily accepted.

But most of all, it is also necessary to consider the other synsets in the net, namely to achieve a balanced encoding of the lexicalized concepts. This way, it seems that synsets such as {rastejar}$_V$ (≅ crawl), {andar}$_V$ (≅walk), {deslizar}$_V$ (≅slide), and so on, establish a similar relation with the concept denoted by {mover-se}$_V$ (≅ move oneself/itself) – they all denote concepts that specify a particular manner of occurrence of the movement event –, and, not having auto-hyperonyms of their own, are necessarily linked to the synset denoting the general concept. The same occurs with change of position verbs, where verbs such as {espreguiçar-se}$_V$ (≅ stretch out) or {esbracejar}$_V$ (≅move the arms, gesticulate) are hyponyms of {mover-se}$_V$ (≅ move the whole body or parts of it voluntarily), not having corresponding auto-hyperonyms, regardless of the presence of the clitic *–se*.

The option to encode these verbs as hyponyms of synsets denoting a more underspecified concept, instead of as hyponyms of what could be considered their direct auto-hyperonym, results in sets of co-hyponyms that share a similar base concept, grouped under the same node in the net and forming naturally emerging classes, corresponding, for instance, to manner of motion verbs, verbs of directed motion, and so on.

2.3 Conclusion

In this chapter we focused on the construction of the Portuguese wordnet of verbs of movement and on the issues associated to this task, namely the different levels of analysis that are relevant for a coherent encoding of the verbs of this class.

Implemented in WN.PT, verbs of movement are encoded using lexical-semantic similarity relations (SYNONYMY, NEAR SYNONYMY), semantic opposition relations (antonymy, NEAR ANTONYMY), subtyping relations (HYPERONYMY/HYPONYMY), part-whole relations (HAS SUBEVENT/IS SUBEVENT OF, HAS TELIC SUBEVENT/IS TELIC SUBEVENT), cause relations (CAUSES/IS CAUSE OF), role relations (INVOLVED AGENT/ROLE AGENT; INVOLVED PATIENT/ROLE PATIENT, etc.) and cross POS relations encoding associations of different order (IN MANNER/MANNER OF).

The definition of the lexical hierarchy for the Portuguese verbs of movement required the definition of the top nodes of the net as well as of some other encoding options, in order to assure conceptual consistency throughout the net and to test the inheritance of semantic information from the higher to the lower nodes in the hierarchy. This task revealed several issues concerning the lexical conceptual organization of the verbs considered, but also issues related to the semantic and syntactic diversity of directly related verbs within the net. The main options defined concerned the unification of causative and non-causative subnets of verbs of movement, assuming, on the one hand, that the causative and inchoative readings of a given verb do not reflect two distinct lexical entries, and, on the other, that the differences in the argument structure of the predicates considered, namely number of arguments and selection properties, are associated to the semantic properties of lexical items, and directly related to the specification of the concepts denoted.

As underlined in Marrafa (2002), the encoding decisions in relational lexica are not straightforward, since the degree of specification of meaning often depends on the goals pursued. Given our goal to construct a relational lexicon for verbs of movement, providing semantic grounds for the syntactic behavior of the lexical items, it is necessary to further analyze the meaning components that differentiate the verbs encoded in WN.PT. The next chapter is, thus, dedicated to the definition of the semantic elements lexicalized by the hyponym synsets, the analysis of emerging patterns, and the observation of the correlation between the semantic properties and the syntactic behavior of these verbs.

3. Hyponymy and lexicalization patterns

3.0 Introduction

The analysis that leads to the organization of lexical items in a wordnet carries strategies of decomposition of meaning: in order to establish hyperonymy/hyponymy relations between synsets it is necessary to determine the basic concept lexicalized by a given synset, identifying the specific meaning contribution of the synset at stake with respect to its hyperonym. Furthermore, establishing different synsets as hyponyms of a given node denoting a less specific concept allows us to observe what distinguishes co-hyponym nodes, i.e., daughters nodes of the same synset.

Examining co-hyponyms, as well as the analysis of the hierarchical structure of a given subnet, specifically the subnets of direct hyponyms, can lead us to determine the specific meaning components that are added to the base concept denoted by the common hyperonym and if any lexicalization patterns emerge.

The analysis of hyponymy and of semantic incorporation phenomena, i.e., the lexicalization of semantic components that can be individuated, allows us to identify the semantic components that differentiate the nodes in a lexical-conceptual net and the lexical semantic properties that may account for the semantic and syntactic behavior of the verbs in study.

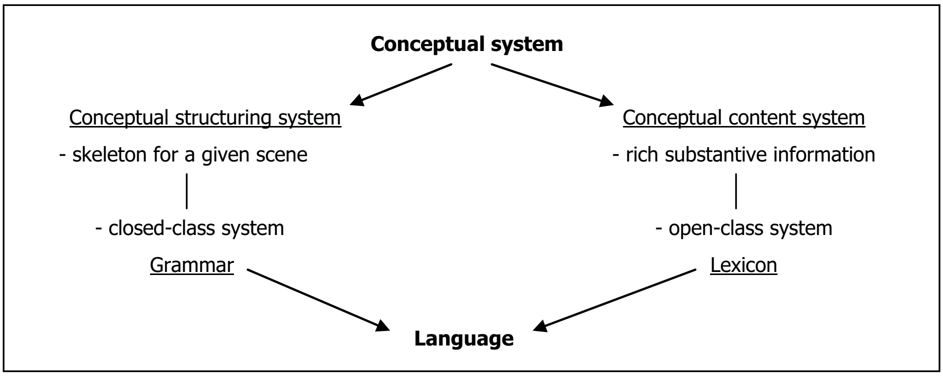
In this chapter we examine an extensional approach to verb meaning (the WordNet model), in which word meaning is represented by relations, with a compositional one (Talmy 1985, Jackendoff 1983, 1990, Rappaport Hovac & Levin 1998) that represents word meaning through semantic decomposition. We focus on the motivations for the observation of semantic incorporation in verbs of movement, and present our analysis of Portuguese data, in what concerns lexicalization patterns, co-hyponym compatibility, and the correlation between semantic lexicalization at hyponym level and verbal argument structure and Aktionsart properties.

3.1 Semantic incorporation and lexicalization patterns

Semantic incorporation, in the sense adopted here, is a common notion in Cognitive Semantics and refers to the process of linguistic expression of concepts or, more specifically, the process of conveying rich substantive information from the conceptual content system into the Lexicon. Cognitive Semantics is concerned with the relationship between experience, cognition and language, and explores the connection between human bodily experiences, the conceptual system and the semantic structure expressed by language (see Evans & Green (2006) for a comprehensive introduction to Cognitive Semantics).

Within Cognitive Semantics, Talmy (2000a) describes language as a system of mental reasoning through which it is possible to study the mind in general: "[R]esearch on cognitive semantics is research on conceptual content and organization in language and, hence, on the nature of conceptual content and organization in general" (Talmy 2000a: 4). The author proposes that the conceptual system is subdivided into the conceptual structuring system (which provides the base structure of a given situation) and the conceptual content system (which provides conceptual information about the situation). These two systems correspond, respectively, to language closed-class system (the Grammar) and language open-class system (the Lexicon), as represented in (1).

(1) Conceptual system and Language in Talmy's theory



Within this theory, Talmy (2000b) focuses on Motion events and on how these events are expressed in natural languages. He considers a Motion event to be an event (portion of reality delimited by the human mind) in which there is a component of Motion: "The component of **Motion** (with capital M) refers to the presence *per se* of motion or locatedness in the event" (Talmy 2000b:25).

A basic Motion event has four components – Motion, Figure, Ground and Path – and consists of the **Figure** (moving or conceptually moveable entity) in **Motion** (moving or being located) with respect to the **Ground** (stationary object or frame that establishes a reference with respect to which the path or location of the Figure is characterized) through or in a given **Path** ("path followed or site occupied by the Figure object with respect to the Ground object" (Talmy 2000b: 25). A basic Motion event is also frequently associated with an external **Co-event** "that most often bears the relation of Manner or of Cause to it." (Talmy 2000b:25).

Following this analysis, Talmy (1985, 2000b) studies how natural languages express motion events, i.e. how natural languages express the five semantic components established, and determines the relevant patterns encountered.

3.1.1 Talmy's typologies

Talmy proposes two typologies (Talmy 1985, 1991, 2000b) according to two different questions considered. In a first perspective, natural languages are classified according to the morphosyntactic constituents that characteristically express Path, originating a two-way typology of verb-framed languages and satellite-framed languages. On another approach, natural languages are classified according to the semantic component lexicalized in the verb, establishing a three-way typology of Motion+Co-event pattern languages, Motion+Path pattern languages and Motion+Figure pattern languages.

3.1.1.1 Path encoding typology

In what concerns the morphosyntactic constituents that characteristically express the Path component in a Motion event, considering that the verb encodes Motion, it is possible to have:

i) Verb-framed languages, such as Romance languages, Semitic languages, Turkic languages, Basque, Japanese and Korean, where the Path is typically lexicalized in the verb, as exemplified by the sentence in (2).

(2) a. La botella *salió* de la cueva. (example from Férez 2008:37)
(≅ the bottle moved out of the cave)

ii) Satellite-framed languages, such as Germanic languages, Slavic languages, Finno-Ugric languages, Mandarin and Walpiri, where the Path is typically expressed by satellites ("grammatical category of any constituent that is in a sister relation to the verb" (Talmy 2000b:102)), as exemplified by the sentence in (3).

(3) a. The bottle floated *out* of the cave. (example from Férez 2008:37)

Talmy (1991, 2000b) suggests that this typology applies also to the expression of change of state and action correlation, i.e. in verb-framed languages the verbs typically incorporate change of state and action correlation (in (4)a and b and (5)a and b), whereas in satellite-framed languages the change of state and action correlation are typically expressed by satellites (in (4)c and d and (5)c and d):

(4) Change of state expression

a. *Apagué* la vela de um soplido. (≅ I put the candle out with a blow)

b. Lo *maté* con fuego/quémandolo. (≅ I killed him with fire/by burning him)

c. I blew *out* the candle.

d. I burned him *to death*.

(5) Action Correlation expression

a. Yo lo *acompañé* tocando la melodía. (≅ I accompanied him playing the melody)

b. Yo lo *superé* tocando la melodía. (≅ I outdid him playing the melody)

c. I played the melody *along* with him.

d. I *out*played him.

(examples adapted from Férez (2008:39))

Several authors[1] have noticed that, although referring the more frequent or prominent patterns, Talmy's typology does not apply adequately to some languages. On the one hand, there seems to be significant differences among languages in the same typological group, as is the case of Spanish and Basque, for instance, in what concerns the expression of Path and Manner in Motion events (Ibarretxe-Antuñano 2004). On the other hand, there are some languages that do not exhibit a prominent pattern for Path expression, using both satellites and verb to express Path in Motion events, which motivated the consideration of a third class of languages, the equipollently-framed languages (see Slobin (2004)).

Our observation of Portuguese verbs shows that, if it is true that Path expressing verbs seem to be more frequent in Portuguese than in English, for instance, it is also true that satellites expressing Path are also very frequent in Portuguese, especially in sentences where the main verb is a manner of motion verb:

(6) a. A massa **flutuou** <u>para o interior do</u> disco, o Sol formou-se no centro e o momento angular foi transferido para o exterior, de tal forma que agora reside principalmente nos planetas. (The mass **floated** <u>into</u> the disc, the Sun formed in the center and the angular moment was transferred to the exterior, in such a way that it now resides mainly in the planets) (http://w3.ualg.pt/~jluis/cap_2.pdf)

[1] See Slobin (2004), Zlatev & Yangklang (2004), Ibarretxe-Antuñano (2004), for instance.

b. Se o barco estiver ancorado, então o peixe tenderá a ser mostrado no visor como linhas horizontais à medida que eles **nadam** para dentro e para fora do feixe do transdutor do sonar. (If the boat is anchored, then the fish will appear on the screen in the form of horizontal lines as it **swims** in and out of the sonar transducer beam) (http://www.navmanmarine.net/upload/Marine/Internet_Manuals/4350_4380/4350_4380_MN000240B_por_web.pdf)

c. Mais baratos que os ryokans são os hotéis-cápsula, nos quais os hóspedes **rastejam** para dentro de contentores prefabricados com apenas um colchão e uma televisão lá dentro. (Cheaper than ryokans are capsule-hotels, in which guests **crawl** into prefabricated containers with only a mattress and a television inside)(http://www.travelgate.pt/index.php?option=com_content&view=section&id=108&Itemid=238)

d. Pedaços de borboleta não costumam ir parar em livros. De mariposas, sim; porque elas **voam** para dentro de casa, onde há livros. (Pieces of butterflies don't usually end up in books. Of moths, yes: because they **fly** into the house, where there are books) (http://www.scribd.com/doc/3992400/Geraldine-Brooks-As-Memrias-do-Livropdfrev)

Also, regarding the expression of change of state, sentences such as the one in (4)d are quite productive in Portuguese, as showed by the examples in (7).

(7) a. Noivo bebe até à morte em casamento (Groom drinks to death in wedding) (http://www.observatoriodoalgarve.com/cna/noticias_ver.asp?noticia=29642)

b. Trabalhadores chineses espacam director até à morte (Chinese workers beat director to death) (http://economico.sapo.pt/noticias/trabalhadores-chineses-espancam-director-ate-a-morte_66051.html)

c. CrimRui e Ricardo, de 17 e 18 anos, são suspeitos de torturar até à morte António Mota. (CrimRui and Ricardo, of 17 and 18 years age, are suspects of torturing António Mota to death) (http://www.gforum.tv/board/1513/299123/suspeitos-libertados-torturaram-ate-morte.html)

c. Duas mulheres foram apedrejadas até à morte numa prisão iraniana. (Two women were stoned to death in an Iranian prison) (http://penademorte.planetaclix.pt/ARQ008.htm)

In what concerns the goals in pursue in the present work, although Portuguese does present verbs expressing Path in a regular fashion, as most Romance languages do, ruling out other lexicalization patterns for the Path component does not conform to the data. As it will be shown further ahead, in order to account for semantic lexicalization in Portuguese verbs of movement, it is necessary to consider more than the Path component.

3.1.1.2 Semantic components in verbs typology

With respect to the semantic components that are characteristically lexicalized in verbs, Talmy presents a three-way typology, mirroring the most prominent lexicalization patterns. The three salient patterns are:

i) Motion+Co-Event (Cause or Manner) languages, such as English, where Manner or Cause components are lexicalized in the verb, as exemplified in (8).

(8) a. The barrel slid. (+Manner, non-agentive)

b. I slid. (+Manner, agentive)

c. The paper blew. (+Cause, non-agentive)

d. I pushed the barrel (+Cause, agentive).

ii) Motion+Path languages, such as Spanish, where the Path component, as seen previously, is lexicalized in the verb, as showed by the sentences in (9).

(9) a. La botella entró a la cueva (≅The bottle entered the cave) (+ Path, non-agentive)

b. Meti la botella en la bodega (≅ I inserted the bottle in the cellar) (+Path, agentive)

iii) Motion+Figure languages, such as Navajo and Hokan languages, where the verb lexicalizes Motion and Figure, i.e., the object that changes location. The following sentences exemplify the few cases in which it is possible to consider the semantic incorporation of Figure in languages like Portuguese ((10)a and b) or English ((10)a' and b').

(10) a. Choveu dentro do quarto. (+Figure, non-agentive)
a'. It rained into the bedroom.

Once again, this typology of general patterns does not account for all the data, although there seems to be a significantly larger number of verbs of movement expressing Manner in English, for instance, than there is in other languages such as Portuguese or Spanish. However, and as noticed by Gutiérrez (2001) or Ibarretxe-Antuñano (2004), for instance, Motion+Manner lexicalization is a rather frequent pattern in Spanish and in Basque. In fact, the analysis of Portuguese verbs of movement shows that the lexicalization of Motion+Manner in verbs is a productive process, as illustrated in the sentences in (6), above, and as we will see in the remainder of this chapter.

Although the classification of verbs in terms of Path expression or of the semantic components expressed is not really pertinent to the case at hands, Talmy's analysis of verbs of movement, considering the semantic components lexicalized in verbal items, constitutes the base grounding for the analysis of Portuguese verbs of movement we focus on this chapter. The identification of the semantic components lexicalized in verbs can provide us with the semantic bases for co-hyponym differentiation, as well as for the determination of their semantic and syntactic behavior.

The next section is, thus, dedicated to the analysis of Portuguese verbs of movement, in what concerns the lexicalization of semantic components, considering both the major and the minor patterns identified by Talmy (1985, 2000b), but also considering a new set of semantic components motivated by the observation of the data.

3.2 Lexicalization in a wordnet of Portuguese verbs of movement

The organization of lexical items in a wordnet requires the analysis of the base concept denoted by a given lexical item but also of the specific meaning contribution that distinguishes it from its hyperonyms. The hyponym relation test shows how this is accomplished:

(11) V_1 is hyponym of V_2 iff

a) to V_1 is to V_2+$AdjP_i$/$AdvP_j$/PP_k/NP_l but to V_2 is not to V_1+$AdjP_i$/$AdvP_j$/PP_k/NP_l,

b) He/It V_2-ed but he/it did not V_1

a. *andar* (walk) is *mover-se* (move oneself) by placing the locomotion limbs one after the other, being the body weight supported by the limbs in contact with the ground, but *mover-se* (move oneself) is not *andar* (walk) by placing the locomotion limbs one after the other, being the body weight supported by the limbs in contact with the ground

b. Ele *moveu-se* mas não *andou*. (He *moved* but he did not *walk*.)

andar (walk) IS HYPONYM OF *mover-se* (move oneself)

The determination of the hyponymy relation, in (11), allows us to separate the meaning of the hyponym into components, since the core meaning of the hyponym corresponds to the concept denoted by the hyperonym, the specific meaning of the hyponym being this way identified, as is illustrated in (12), below.

(12) andar(e_1, z):
e_1 = move oneself,
z = placing the locomotion limbs one after the other, being the body weight supported by the limbs in contact with the ground

This way, we can use the hyponymy relation to identify the meaning specificities that distinguish hyponyms from their hyperonyms and to determine the semantic components these meaning specificities correspond to.

In the example presented above, the semantic component lexicalized in the synset $\{andar\}_V$ (walk) (specifying translatory motion) is not entirely consistent with Talmy's (1985) Manner definition, i.e., "a subsidiary action or state that a Patient manifests concurrently with its main action or state" (Talmy 1985: 128), since the way in which the Figure moves (placing the locomotion limbs one after the other, being the body weight supported by the limbs in contact with the ground) does not seem concurrent with the main action of moving, but complementary. Here we adopt Croft *at. al* (2008) reformulation, assuming that MANNER refers to the manner of motion by which the Figure moves along the Path, being the Motion component recovered from the concept denoted by the hyperonym $\{mover\text{-}se\}_V$ (move oneself).

Within a strict WN context – and to determine the base relation of verbal hyponymy – it is not necessary to consider or distinguish between different types of manner (Fellbaum 1998):

(13) *To V1 is to V2 in some particular manner.*
(Fellbaum 1998: 79)

However, as showed by Talmy (1985, 2000), and as confirmed by Portuguese data, these different particular *manners* can refer to different semantic components of a Motion event: Manner, Cause, Path, and so on and. Also, these can be reflected in the argument structure of verbs (number and/or type of arguments selected, as exemplified in (14)) as well as be related to co-hyponym co-occurrence restrictions, in (15), motivating further analyses of verbal hyponymy.

(14) a. *tirar* (take) is *mover* (move) from a given location
$\{tirar\}_V$ is HYPONYM of $\{mover\}_V$

b. Ele moveu <u>o caixote</u>. (He moved the box)

c. Ele tirou <u>o caixote</u> <u>da rua</u>. (He took <u>the box</u> <u>from the street</u>)

(15) a. Ele foi para casa andando. /Ele andou indo para casa
(He went home walking/He walked going home)

b. #Ele andou correndo. /#Ele correu andando.
(He walked running/He ran walking)

c. #Ele foi para casa vindo de casa./#Ele veio de casa indo para casa.
(He went home coming from home/He came from home going home)

The analysis of Portuguese verbs of movement shows that the five semantic components proposed in Talmy (1985, 2000), namely Figure, Ground, Path, Cause and Manner, do not allow a comprehensive treatment of all the verbs of this class. Based on the observation of Portuguese verbs of movement integrated in a wordnet, we propose a new set of semantic components that allow us to characterize hyperonym/hyponym meaning specificities, as well as to account for co-hyponym compatibilities.

3.2.1 New set of semantic components

The set of semantic components proposed here, defined as the meaning elements that correspond to the part of verb meaning that allows us to differentiate it from its hyperonym, results from the study of Portuguese verbs of movement integrated in a wordnet.

(16) Semantic components incorporated in verbal synsets

a. **MANNER**: how the event develops

b. **CAUSE**: what brings about the event

c. **INTENTION**: purpose/intended goal of the event

d. **FIGURE**: object that anchors the event

e. **GROUND**: external object with respect to which the event is put in perspective

f. **SOURCE**: initial location/position of the FIGURE

g. **GOAL**: final location/position of the FIGURE

h. **PATH**: medium locations between the SOURCE and the GOAL

i. **DIRECTION**: way in which the motion event occurs

As previously stated, this enlarged set of semantic components results from the observation of Portuguese verbs of movement. However, the first five components listed in (17)a – (17)e also appear in verbs from other semantic fields, which motivates the reformulation of the original definitions of FIGURE and GROUND components making no specific reference to motion events.

(17) a. +MANNER: *gritar* (shout) is to speak loudly.

b. +CAUSE: *bronzear-se* (tan) is to become dark due to ultraviolet rays exposure.

c. +INTENTION: *sacrificar* (sacrifice) is to kill to please/honor divine entities.

d. +FIGURE: *beber*$_2$ (drink) is to drink alcoholic beverages.[2]

e. +GROUND: *aumentar* (to increase) is to make bigger with respect to a previous dimension.

The MANNER component considered in this set of semantic components excludes all the other considered, thus being much less comprehensive than the manner concept used in the WordNet model.

The CAUSE component proposed here does not correspond to the one proposed by Talmy, namely in what concerns the causative meaning types considered by the author (Talmy 1985: 79). We argue that there is CAUSE lexicalization only in the cases in which the meaning contribution of the hyponym refers to what brings about the event, as the example in (17)b illustrates.

The introduction of the component INTENTION follows the proposal described in Fellbaum (1998a: 80), for communication verbs, and in Barreto (2002: 57), and can be related to the concept of Telic qualia role in the Generative Lexicon, defined as the aspect of meaning of a word that specifies its purpose and function (Pustejovsky 1995: 76), as exemplified in (17)c.

The Path component, originally defined by Talmy as the component responsible for conveying information about the trajectory described by the Figure, is subdivided here. Several authors adopt more specific concepts included in Talmy's notion of Path, such as SOURCE, PATH (set of locations between SOURCE and GOAL), DIRECTION and GOAL.

These distinctions are shown to be relevant for the analysis of verbs like *subir* (move up) and *retroceder* (move back through the same path) that lexicalize the components DIRECTION and PATH. Our data reveals that, in the subset of change of location verbs, DIRECTION lexicalization is the more frequent pattern (*afastar* (move away), *avançar* (move forward), *trazer* (move in our direction), *descer* (move down), *entrar* (move inside), etc.), whereas the lexicalization of GOAL (*pôr* (move to a given final location), *carregar* (put in a transport vehicle), *encaixotar* (put inside a box), *enterrar* (put under ground)), PATH (*enfiar* (put through a narrow opening), *retroceder* (*move back through the same path*), *circum-navegar* (navigate around something), etc.), and

[2] The verb *beber*$_2$ (drink$_2$) refers to the autohyponym verb of the more general concept denoted by the verb *beber*$_1$ (drink$_1$): ingesting liquids.

SOURCE (*tirar* (move from a given location), *remover* (move from the usual location), *desencaixotar* (take from a box) are not as common in the set of verbs observed[3].

Also, SOURCE, GOAL and PATH components, in Portuguese as in English, can be expressed by prepositional phrases, as the one in (18), in cases where the verb already conveys DIRECTION, for instance, presenting evidence that Talmy's notion of Path is, perhaps, too broad.

(18)	O João subiu	do 1º andar	para o sotão	pelas escadas.
	(John went up	from the 1st floor	to the attic	through the stairs.)
	+DIRECTION	SOURCE	GOAL	PATH

The lexicalization of these semantic components is observed at each level of the hierarchy, i.e., we only consider the semantic components lexicalized (+) with respect to the meaning of the hyperonym. In order to systematize this analysis, we developed a battery of formulae to test the semantic component lexicalized in the hyponym.

(19) a. **MANNER**: how the event develops.

Test 1: V_1 is hyperonym of V_2 and

V_2 é V_1 **como**? (to V_2 is to V_1 **how**?)

Ex: *arrastar* é *mover* como? Com esforço em contacto com uma superfície que oferece oposição significativa ao movimento.

(*drag* is *move* how? With effort, in contact with a surface that offers significant resistance to movement)

Test 2: He/it V_2 (something/somenone), but he/it did not V_1+MANNER (something/someone): **False**

Ex: #Ele arrastou o objecto mas não moveu o objecto com esforço em contacto com uma superfície que oferece oposição significativa ao movimento.

(He dragged the object, but he did not move the object with effort, in contact with a surface that offers significant resistance to movement)

[3] Besides Lakoff (1987), Johnson (1987) and Fillmore et al.(2000), already mentioned in the introduction of this work (section 1.2.2.1), Levin (1993), that proposes verb classes considering the semantic notions of direction (verbs of putting with a specified direction, verbs of inherently directed motion) and spatial configuration (verbs of putting in a spatial configuration, verbs of spatial configuration); Clark & Carpenter (1994) examines the notion of *source* in language acquisition, distinguishing *source*, from *direction*, or *goal*; Asher & Sablayrolles (1996) proposes a typology for motion verbs and spatial PPs in French, establishing several types of spatial relations according to the directions of the movement; Gutierrez (2001) argues that the Ground objects may refer to some subpart, as the origin or the endpoint, of the Path; Teixeira (2001) on the verbalization of space; among others.

b. **CAUSE**: what brings about the event.

Test 1: V_1 is hyperonym of V_2 and

V_2 é V_1 **devido a quê**? (to V_2 is to V_1 **due to what**?)

Ex.: *bronzear-se* é *escurecer* devido a quê? Devido à exposição a raios ultravioletas.

(*tan* is *become dark* due to what? Due to the exposure to ultraviolet rays)

Test 2: He/it V_2 (something/somenone), but he/it did not V_1+CAUSE (something/someone): **False**

Ex: #Ele bronzeou-se mas não escureceu devido à exposição a raios ultravioletas.

(He tanned but he did not become dark due to the exposure to ultraviolet rays)

c. **INTENTION**: purpose/intended goal of the event.

Test 1: V_1 is hyperonym of V_2 and

V_1 é V_2 **para quê/com que finalidade**? (to V_1 is to V_2 **for what/with what intention**?)

Ex.: *embalar* é *balançar* para quê/com que finalidade? Para adormecer.

(*rock* is *sway* for what/with what intention? To get someone to sleep)

Test 2: He/it V_2 (something/someone), but he/it did not V_1+INTENTION (something/someone): **False**

Ex: #Ele embalou a criança mas não balançou a criança para a adormecer.

(He rocked the child but he did not sway the child to get him to sleep)

d. **FIGURE**: object that anchors the event

Test 1: V_1 is hyperonym of V_2 and

V_2é V_1 **quem/o quê**?/ V_2 é **quem/o quê** V_1? (to V_2 is to V_1 **who/what**? V_2 is to **who/what** V_1?)

Ex.: *anestesiar* é *administrar* o quê? Anestesia.

(*anesthetize* is *administrate* what? Anesthesia)

Test 2: He/it V_2/He/it V_2 something/someone but he/it did not V_1+FIGURE/ he/it did not V_1+FIGURE something/someone: **False**

Ex: #Ele anestesiou o paciente mas não administrou anestesia ao paciente.

(He anesthetized the patient but he did not administrate anesthesia to the patient)

e. **GROUND**: external object with respect to which the event is put in perspective

Test 1: V_1 is hyperonym of V_2 and

V_2 é V_1 ***relativamente a quê?*** (to V_2 is to V_1 **with respect to what**?)

Ex.: *adiantar-se* é *avançar* relativamente a quê? A algo em movimento.

(*move ahead of* is *move forward* with respect to what? With respect to something in motion)

Test 2: He/it V_2 (something/someone), but he/it did not V_1+GROUND (something/someone): **False**

Ex: #Ele adiantou-se ao colega mas não avançou em relação ao colega.

(He moved ahead of the colleague but he did not move forward with respect to the colleague)

f. **SOURCE**: initial location/position of the FIGURE.

Test 1: V_1 is hyperonym of V_2 and

V_2 é V_1 ***de onde/de que posição?*** (to V_2 is to V_1 **from where/from which position**?)

Ex.: *desencaixotar* é *retirar* de onde? De um caixote.

(*unbox* is *remove* from where? From a box)

Test 2: He/it V_2 (something/someone), but he/it did not V_1+SOURCE (something/someone): **False**

Ex: #Ele desencaixotou a mesa mas não retirou a mesa do caixote.

(He unboxed the table but he did not remove the table from the box)

g. **GOAL**: final location/position of the FIGURE.

Test 1: V_1 is hyperonym of V_2 and

V_2 é V_1 ***onde/para onde/ para que posição?*** (to V_2 is to V_1 ***where/*to where/to which position**?)

Ex.: *encaixotar* é *meter* onde? Num caixote.

(*box* is *put inside* where? In a box)

Test 2: He/it V_2 (something/someone) but he/it did not V_1+GOAL (something/someone): **False**

Ex: #Ele encaixotou a mesa mas não meteu a mesa num caixote.

(He boxed the table but he did not put the table inside a box)

h. **PATH**: medium locations between the SOURCE and the GOAL.

Test 1: V_1 is hyperonym of V_2 and

V_2 é V_1 ***por onde?*** (to V_2 is to V_1 **through where**?)

Ex.: *circundar* é *mover-se* por onde? Em redor de algo.

(*circumbulate* is *move* through where? Around something)

Test 2: He/it V_2 (something/someone), but he/it did not V_1+PATH (something/someone): **False**

Ex: #Ele circundou a estátua mas não se moveu em redor da estátua.

(He circumbulated the statue but he did not moved around the statue)

i. **DIRECTION**: way in which the motion event occurs.

Test1: Test 1: V_1 is hyperonym of V_2 and

V_2 é V_1 ***em que direcção?*** (to V_2 is to V_1 **in which direction**?)

Ex.: *recuar* é *mover-se* em que direcção? Para trás.

(*move back* is *move* in which direction? Backwards)

Test 2: He/it V_2 (something/someone), but he/it did not V_1+DIRECTION (something/someone): **False**

Ex: #Ele recuou mas não se moveu para trás.

(He moved back but he did not move backwards)

As the tests and examples above show, the GROUND component considered here is not completely coincident with the one presented in Talmy (1985). The author describes the verb *box* as incorporating Ground, not GOAL, as proposed by us. Our view is that it is more intuitive to consider that 'box' is the final location of the FIGURE than the object with respect to which the FIGURE is moved, and the tests seem to confirm it:

(20) a. To *box* is to *put* with respect to what? ?#With respect to a box.

?He boxed the table but he did not put the table with respect to a box.

b. To *box* is to *put* where? In a box.

#He boxed the table but he did not put the table in a box.: false

For perspicuity purposes, the examples above present the lexicalization of only one semantic component, but frequently more than one component is lexicalized. The following examples illustrate this phenomenon.

(21) a. +GROUND & MANNER

voar (fly) = move in the air$_{\text{GROUND}}$ with wings or artificial means of propulsion$_{\text{MANNER}}$

b. +FIGURE & MANNER

bombear (pump) = take liquids$_{\text{FIGURE}}$ with a pump$_{\text{MANNER}}$

c. +FIGURE & MANNER & PATH & GROUND

orbitar (orbit) = move, celestial bodies$_{\text{FIGURE}}$, cyclically$_{\text{MANNER}}$, around$_{\text{PATH}}$ another celestial body$_{\text{GROUND}}$

Note that we are only considering semantic lexicalization at each level of the hierarchy: the semantic components identified correspond only to those incorporated by the hyponym with respect to the hyperonym meaning. Although verb wordnets do not present as many hierarchy levels as nominal ones, it is possible to have some subnets three or more levels deep. In these cases, the lower node in the net inherits the semantic content of all its hyperonym nodes, although only the semantic components that are specific to it are considered at that level (cf. (22)a and (22)b below).

(22) a. Hyponymy subtree

{mexer}v + *MOTION*
(*change position*)

- **{abanar}v** +*MANNER & SOURCE & GOAL*
  (**mexer** from one position to the opposite one with impulses in opposite directions; ≃ *shake*)
  - **{balançar, balancear}v** +*MANNER*
    (**abanar** smoothly; ≃ *sway*)
    - **{embalar}v** +*FIGURE & INTENTION*
      (**balançar**, animated entities, to get them to sleep; ≃ *rock, cradle*)

b.

{embalar}v +*MOTION & MANNER & SOURCE & GOAL &FIGURE & INTENTION*
(change the position of animated entities, from one position to the opposite one, with smooth impulses in opposite directions, to get them to sleep; ≃ *rock, cradle*)

3.2.2 Degrees of semantic incorporation

The analysis of semantic incorporation in a wordnet of Portuguese verbs of movement allowed us to observe that the incorporation of semantic components – reflecting the specific meaning of hyponyms – does not always result in a complete lexicalization process, that is, verbs can fully incorporate the semantic components considered – lexicalization – or incorporate semantic restrictions referring to the semantic components considered. In the last case, the incorporation of the semantic restrictions referring to the semantic components considered often results in changes in argument structure and argument selection specifications, as observed by Fellbaum (1998a): the deeper the level of hyponymy, the greater the number of hyponyms, and, consequently the stricter are the selection constraints of the verbs, directly related to the specificity of the concepts denoted, affecting the syntactic expression of predicates to different degrees.

The fragment of the constructed wordnet in (23) illustrates this phenomenon, considering only the basic adicity of the verbs, i.e. the overt arguments that are necessarily realized.

(23) Hyponym subtree with information on the overt arguments of the verbs

{mover, deslocar}v [change location]
V(ARG$_1$, ARG$_2$)
- **{pôr, colocar}v** [*mover* to a given final location; ≅ *put*] **+GOAL**
  V(ARG$_1$, ARG$_2$, ARG$_3$)
  - **{meter}v** [*pôr* inside of; ≅ *put inside of*] **+DIRECTION**
    V(ARG$_1$, ARG$_2$, ARG$_3$)
    - **{enjaular}v** [*meter* in a cage; ≅ *cage*) **+GOAL**
      V(ARG$_1$, ARG$_2$)

The verbs in the hyponym synset {*pôr, colocar*}$_V$ (≅ put) are semantically stronger than their hyperonym, given that they incorporate semantic restrictions on the GOAL component, thus presenting a larger list of arguments, in comparison with the argument structure of the hyperonym. The expression of the final location, not accounted for in the argument structure of the hyperonym, becomes mandatory. The direct hyponym, *meter* (≅ put inside of), denotes a specific DIRECTION of the motion event but presents the same number of overt arguments as its hyperonym[4]. The lowest hyponym, *enjaular* (≅ cage), in turn, lexicalizes a GOAL component, *cage*, resulting in the reduction of the number of overt arguments of the verb, in a process similar to *lexical shadowing,* described in Pustejovsky (1995, 2000).

The semantic component whose lexicalization is less common is FIGURE. Besides a small group of verbs, such as *chover* (rain), *anestesiar* (anaesthetize), *beber* (drink, consume alcoholic beverages regularly), *fumar* (smoke, smoke cigarettes) or *mover-se* (move oneself), the majority of which constitute auto-hyponyms (pairs of synsets in which the lexicalization of an argument results in a simultaneously homonym and hyponym verb with a different meaning) that lexicalize one of the overt arguments of the hyperonym, most verbs whose meaning refers to FIGURE only exhibit restrictions on the argument that corresponds to the object that anchors the event. Below, we list the Portuguese verbs of movement that incorporate semantic restrictions on FIGURE, isolated or with other semantic component.

[4] As we will discuss further ahead, the lexicalization or the incorporation of semantic restrictions of the semantic components considered, always results in the increase of the argument structure of the hyponym verbs, for the reasons explained above. However, these arguments may or not correspond to overt arguments.

(24) FIGURE in Portuguese verbs of movement

a. {drenar 1}$_V$ [≅ drain; take out liquids from soil] (+SOURCE)

b. {entornar, derramar}$_V$ [≅ overturn, spill; take out of, typically liquids]

c. {trasladar}$_V$ [≅move human remains to another grave] (+GOAL)

d. {bombear}$_V$ [≅ pump; take out liquids using a pump] (+MANNER)

e. {escoar}$_V$ [≅ drain; take out liquids, using channels and openings] (+MANNER)

f. {arrancar}$_V$ [extract, objects with roots, by the root] (+MANNER)

g. {debandar}$_V$ [≅ disband; run away, groups of entities, in a disorganized and hurried way] (+MANNER)

h. {galopar}$_V$ [≅ gallop; run, equines and ruminants, performing three-tempo jumps] (+MANNER)

i. {trotar}$_V$ [≅ trot; run, equines and ruminants, placing two of the four limbs on the ground, at a time] (+MANNER)

j. {drenar 2}$_V$ [≅ drain; take out liquids from organic cavities using a drain.] (+MANNER & SOURCE)

k. {retirar}$_V$ [≅ retreat; exit, troops, from the battle field] (+SOURCE)

l. {correr 2}$_V$ [≅ run; move, liquids, by gravity] (+MANNER)

m. {montar}$_V$ [≅ mount; place anthropomorphic entities on top of something with one leg on each side of it] (+MANNER)

n. {orbitar}$_V$ [≅ orbit; move, celestial bodies, cyclically, around another celestial body] (+MANNER & PATH & GROUND)

o. {baloiçar, balouçar}$_V$ [swing, at pace, something suspended] (+MANNER)

p. {dispor}$_V$ [≅ arrange, order; put objects of a set according to a specific order] (+MANNER)

q. {sentar}$_V$ [≅ sit; place anthropomorphic entities in a position in which the body weight is supported by the buttocks] (+MANNER)

r. {bater}$_V$ [≅ beat; move, the heart, at pace due to involuntary muscular contractions and dilations] (+MANNER & CAUSE)

s. {revolver, remexer}$_V$ [≅ rummage; move the objects in a given space, changing their relative position] (+MANNER & GROUND)

t. {acocorar-se}$_V$ [≅ squat; place oneself in a position in which the bodyweight is supported by the feet and the torso is parallel to the thighs, by bending the knees] (+GOAL & MANNER)

u. {ajoelhar-se}$_V$ [≅ kneal; place oneself in a position in which the bodyweight is supported by the knees] (+GOAL)

v. {sentar-se}$_V$ [≅ sit; place oneself in a position in which the body weight is supported by the buttocks] (+GOAL)

w. {empinar-se}$_V$ [≅ rear; place oneself, quadrupeds, with the anterior feet off the ground and held up the torso in a vertical position] (+GOAL)

With the exception of the verb in (24)a, where there is in fact lexicalization of the FIGURE, all the other verbs listed select an argument expressing FIGURE that conforms to the semantic type restrictions imposed by the verb semantic properties, as the examples in (25) demonstrate.

(25) a. As éguas galoparam (pelo campo). (The mares galloped (through the field))
a'. #Os patos galoparam (pelo campo). (The ducks galloped (through the field))

These two degrees of semantic incorporation – lexicalization of a given semantic component and incorporation of restrictions on a given semantic component – were both considered in our analysis of the data. For purposes of quantification, we do not divide the cases of lexicalization from those of incorporation of restrictions on the semantic components. However, it is clear that the determination of the semantic components lexicalized sheds some light on the differences concerning the syntactic realization of lexical-conceptually related verbs.

In the next section, we will present the quantitative results concerning the occurrence of semantic components determined in the wordnet of Portuguese verbs of movement developed, as well as the analysis of hyponyms in what concerns co-hyponym compatibility, and differences in argument structure and Aktionsart properties of hyponym verbs with respect to their hyperonym.

3.2.3 Portuguese data: quantitative analysis

The wordnet of Portuguese verbs of movement built under the scope of this work amounts to a total of 214 synsets. The subnet of change of location verbs is constituted by 121 synsets; that of change of position verbs includes 93 synsets. The following table presents the quantitative data regarding the lexicalization of the semantic components previously considered in the hyponym synsets of the top nodes {mover, deslocar}$_V$ (≅ change location) and {mexer}$_V$ (≅ change position), in isolation or with other components.

Change of location verbs {mover, deslocar}$_V$ (+MOTION)	
Semantic components	**Total**
+MANNER	36
+DIRECTION	17
+GOAL	17
+SOURCE	8
+PATH	7
+FIGURE	2
+GROUND	2
+INTENTION	1
+CAUSE	0
+MANNER & FIGURE	9
+MANNER & GROUND	5
+MANNER & PATH	3
+GOAL & SOURCE	2
+FIGURE & SOURCE	2
+DIRECTION & SOURCE	1
+DIRECTION & GROUND	1
+GOAL & DIRECTION	1
+GOAL & FIGURE	1
+MANNER & DIRECTION	1
+MANNER & GOAL	1
+MANNER & INTENTION	1
+MANNER & SOURCE	1
+MANNER & FIGURE & SOURCE	1
+MANNER & PATH & FIGURE & GROUND	1
	Total 121

Change of position verbs {mexer}$_V$ (+MOTION)	
Semantic components	**Total**
+GOAL	28
+MANNER	21
+INTENTION	2
+CAUSE	1
+DIRECTION	2
+FIGURE	1
+GROUND	0
+PATH	0
+SOURCE	1
+MANNER & GOAL	6
+GOAL & FIGURE	6
+GOAL & DIRECTION	6
+GOAL & GROUND	4
+MANNER & CAUSE	3
+MANNER & FIGURE	3
+MANNER & DIRECTION	2
+MANNER & GROUND	2
+MANNER & INTENTION	2
+MANNER & FIGURE & CAUSE	1
+MANNER & FIGURE & GROUND	1
+MANNER & GOAL & SOURCE	2
	Total 93

Table 1: Semantic components lexicalized in Portuguese verbs of movement

Regarding lexicalization patterns in Portuguese verbs of movement, i.e., the frequency and forms of lexicalization or incorporation of restrictions on the semantic components considered, it is possible to verify that the lexicalization of more than one component in hyponym synsets is the most frequent pattern (31,3%), as showed below. Note, also, that these data only refer to the lexicalization of semantic components at hyponym level; lexicalization processes already established in the hyperonym meaning are not considered.

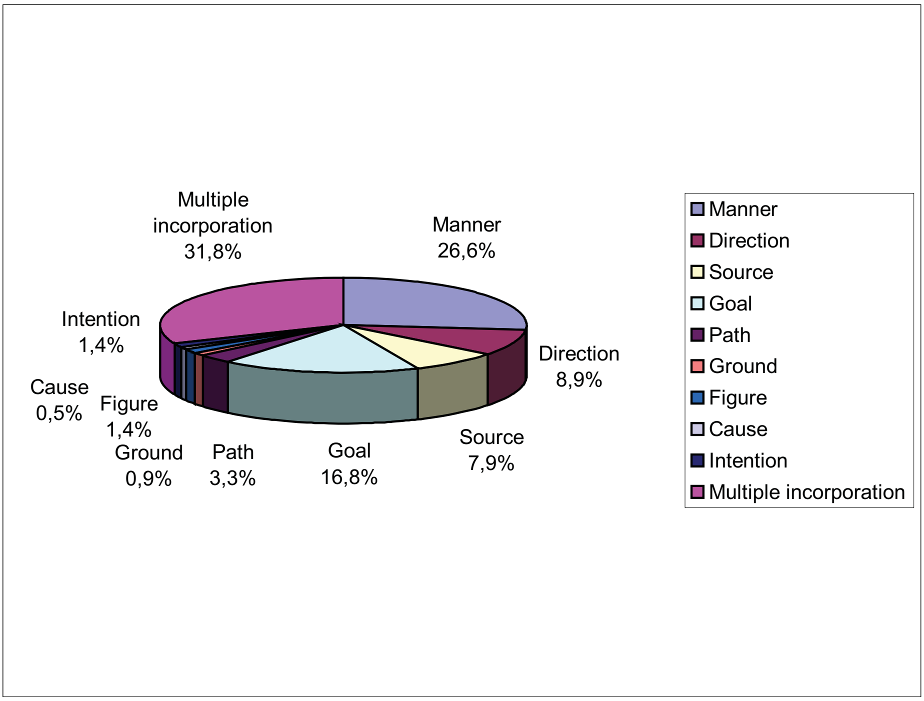


Figure 1: Lexicalization of semantic components in Portuguese verbs of movement

The simultaneous lexicalization of more than one semantic component is the most common case in Portuguese verbs of movement (31,8%). The lexicalization of MANNER, in isolation, is the next most frequent pattern (26,6%). However, the sum of the cases of lexicalization of PATH, GOAL, SOURCE and DIRECTION (which jointly constitute the Path component considered by Talmy (1985, 2000b)) surmounts this percentage, adding up to 36,9% of the cases observed. The lexicalization of the semantic components GROUND, FIGURE, INTENTION and CAUSE in isolation is almost irrelevant (lower than 1,5%).

The data concerning the distribution of the semantic components by the synsets observed present slightly different results. In this perspective, we add up all the cases of lexicalization of a given semantic component, whether in isolation or simultaneously with other semantic components. Figure 2, below, presents this distribution. As can be observed, the lexicalization of MANNER still is the most salient pattern for Portuguese, considering the subdivision of Talmy's (1985, 2000b) Path component. GOAL is the second semantic component more frequently present in Portuguese verbs of movement, followed by DIRECTION and FIGURE components.

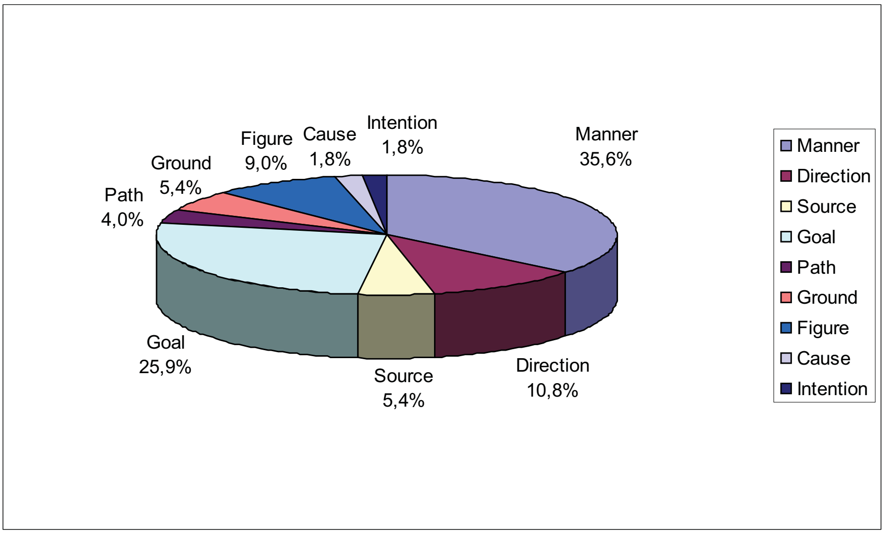


Figure 2: Distribution of semantic components in Portuguese verbs of movement

With regard to the subset of change of location verbs, in Figure 3, MANNER continues to be the most frequent semantic component lexicalized, followed by GOAL and DIRECTION.

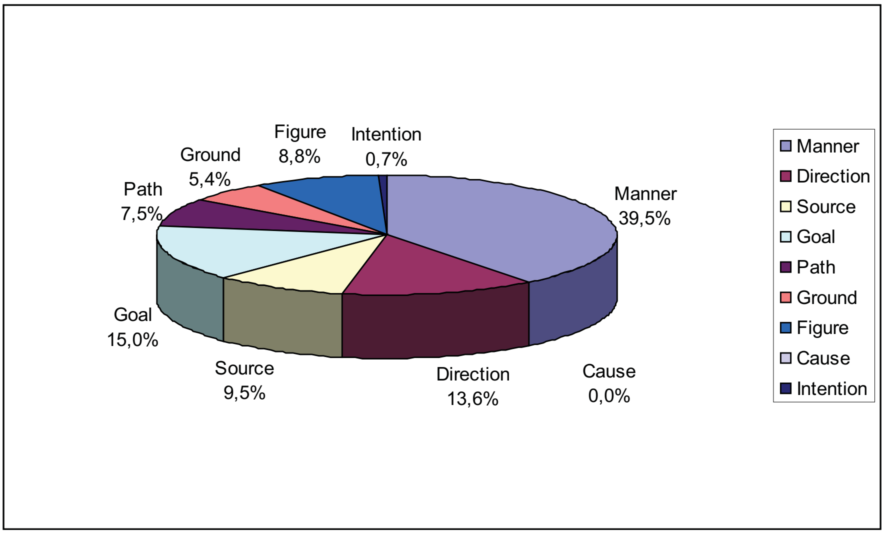


Figure 3: Distribution of semantic components in Portuguese change of location verbs

This distribution is not mirrored in change of position verbs. In this case, GOAL is the most frequent component lexicalized, followed by MANNER, as showed in the Figure 4 below.

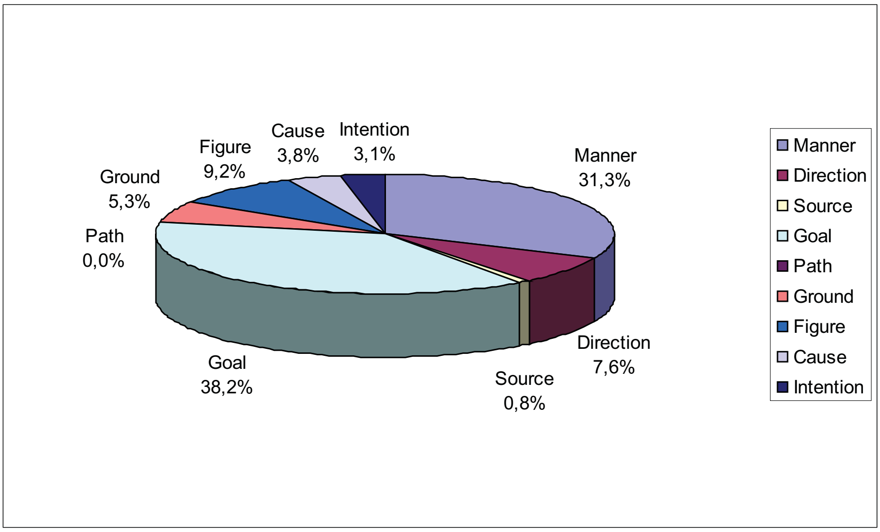


Figure 4: Distribution of semantic components in Portuguese change of position verbs

The lexicalization of GOAL – the final position – in change of position verbs is the most frequent case, i.e., many change of position verbs specify the position in which the entity ends up, disregarding the initial position of the entity: {curvar, dobrar, encurvar, entortar}$_V$ (≅ move into a curve position), {endireitar, desencurvar, desentortar}$_V$ (≅ move into an unfolded position), {amontoar}$_V$ (≅ place in a bundle), {encavalitar}$_V$ (≅ place on top of each other), {enfileirar}$_V$ (≅ place in a row), {acomodar}$_V$ (≅ put in a comfortable or adequate position), {endireitar}$_V$ (≅ put in the correct position), {entortar}$_V$ (≅ put in an incorrect position), {enviesar*,* enviusar}$_V$ (≅ put in a oblique position), {estender}$_V$ (≅ put in a straight, unfolded position), {deitar}$_V$ (≅ put in a horizontal position), {inclinar}$_V$ (≅ put in an oblique position; bend), {virar}$_V$ (≅ put in a different (typically opposite) orientation), {pôr-se}$_V$ (≅ move oneself assuming a position), {apoiar-se}$_V$ (≅ move oneself into a leaning position), {arquear-se}$_V$ (≅ move oneself into an arch position), {curvar-se, dobrar-se}$_V$ (≅ move oneself into a curve position), {deitar-se}$_V$ (≅ move oneself into a horizontal position), {estender-se}$_V$ (≅ move oneself into a straight, unfolded position), {inclinar-se}$_V$ (≅ move oneself into an oblique position), {pendurar-se}$_V$ (≅ move oneself into a suspended position), {voltar-se}$_V$ (≅ move oneself into the opposite position), etc. Also, in the data observed, the SOURCE and PATH components of the movement event are almost never specified in the meaning of the hyponyms.

This quantitative analysis supports our previous observation that the prediction of Talmy's typologies does not conform to the data. MANNER is, in fact, frequently lexicalized in Portuguese verbs of movement, and not only typically in satellites within the sentence. If any, Portuguese should be considered an equipollently-framed language, i.e., a language that expresses Manner and Path both in verbs and sentence satellites (see Slobin 2004).

More important to the current study is the fact that the analysis of the meaning of hyponym verbs in terms of MANNER, CAUSE, INTENTION, FIGURE, GROUND, SOURCE, GOAL, PATH and DIRECTION semantic components can be directly related to co-hyponym compatibility, argument structure differences and Aktionsart properties, as we will show in the next section.

3.2.4 Decompositional analysis of hyponymy

The decompositional analysis of hyponymy in terms of MANNER, CAUSE, INTENTION, FIGURE, GROUND, SOURCE, GOAL, PATH and DIRECTION semantic components lexicalized by Portuguese verbs of movement only considers, as mentioned previously, what occurs at hyponym level, i.e., what distinguishes an hyponym from its hyperonym, assuming that the hyponym inherits all the conceptual properties of its hyperonym. The hyponym is, thus, all that its hyperonym is, plus some additional semantic contribution. For this reason, this analysis aims at providing lexical semantic grounds for explaining different semantic properties of verbs, reflected in different syntactic behaviors, in synsets related by hyponymy. Our observation shows us that these differences concern mainly co-hyponym compatibility, predicate argument structure and Aktionsart properties.

3.2.4.1 Co-hyponym compatibility

As previously shown, the hyponymy relation between synsets reflects the lexicalization of different semantic components, which distinguishes co-hyponyms, i.e. sister synsets (see Fellbaum 1998). According to Mendes & Chaves (2001), this meaning distance is responsible for some incompatibility of nominal co-hyponyms, being co-hyponyms compatible only if they do not lexicalize values for the same semantic feature. As demonstrated by the asymmetry between the sentences in (26)a and b, co-hyponyms such as {fox terrier}$_N$, {police-dog}$_N$ and {pit-bull}$_N$ are not identically compatible.

(26) a. Rex is a fox terrier and a police-dog.

b. #Rex is a fox terrier and a pit-bull.

This contrast results from the fact that {fox terrier}$_N$ and {pit-bull}$_N$ denote the specification of their hyperonym, {dog}$_N$, in what concerns its physical properties. A *fox terrier* is a short, small dog with long curly hair, whereas a *pit-bull* is a medium sized and strongly built dog, with short hair. Naturally, the same dog cannot be simultaneously small and medium sized with long and short hair. On the other hand, a *police-dog* is a dog trained to execute certain police tasks, such as detecting drugs or controlling mobs, but the concept specification does not refer to any physical properties of the dog at stake and thus any *fox terrier* or *pit-bull* or *poodle* can also be *police-dogs*. In other words, {fox terrier}$_N$ and {pit-bull}$_N$, on the one side, and {police dog}$_N$, on the other, exemplify the ontological distinction between "types" and "roles", respectively.

The same type of phenomenon occurs with verbal concepts. The analysis of hyponymy in the wordnet of Portuguese verbs of movement shows that co-hyponyms that lexicalize different semantic components are compatible (see (27)), and that some co-hyponyms lexicalizing the same semantic component are not (see (27)a and c):

(27) {mover-se}$_V$ (≅ move (oneself)) has hyponyms

{contornar}$_V$ (≅ move near the limits of) +PATH
{circundar}$_V$ (≅ circumbulate) +PATH
{rastejar}$_V$ (≅ crawl) +MANNER
{correr}$_V$ (≅ run) +MANNER

a. #O rapaz contornou a estátua circundando-a.
(The boy moved near the limits of the statue circumbulating it)

b. O rapaz contornou a estátua correndo/rastejando.
(The boy moved near the limits of the statue running/crawling)

c. #O rapaz rastejou à volta da estátua correndo.
(The boy crawled around the statue running)

It is possible, for instance, to have two co-hyponym verbs denoting PATH and MANNER in a same sentence, but not co-hyponym verbs both denoting either MANNER or PATH. However, the lexicalization of the same semantic components does not account for the incompatibilities displayed by the data. Let us consider, for instance, the synsets {avançar}$_V$ (≅ move forward) and {descer}$_V$ (≅ move downwards; descend) and {recuar}$_V$ (≅ move backwards), all hyponyms of {mover-se}$_V$ (≅ move oneself) and all incorporating DIRECTION.

(28) a. O rapaz avançou descendo.
(The boy moved forward descending)

b. O rapaz desceu recuando.
(The boy descended moving backwards)

c. #O rapaz avançou recuando.
(The boy moved forward moving backwards)

The examples in (28)a and b show that, in fact, co-hyponym incompatibility results from the lexicalization of opposite or otherwise incompatible values for the same semantic component, and not only from the lexicalization of the same semantic component, as observed by Mendes & Chaves (2001). In the examples presented, the verbs *avançar* (≅ move forward) and *recuar* (≅ move backwards) lexicalize opposite directions of the movement, forward and backwards, respectively, rendering impossible their co-occurrence in the same sentence. The verb *descer*, on the other hand, refers to a different and not incompatible direction of movement, thus being compatible with both *avançar* and *recuar*.

This observation raises the question of how to account for this phenomenon and motivates the need for richer informational structures at the lexical entry level.

3.2.4.2 Argument structure

A first step of this analysis concerned the number of arguments, or adicity, of the predicates. In this section, we will focus on the differences in terms of the number of arguments of verbs with regard to their hyperonyms, and their correlation with lexicalization patterns.

Since the requirement of syntactic realization of arguments is variable, i.e., there are some arguments (shadow arguments, corresponding to lexicalized arguments, for instance) that are not required to be overtly realized, in order to compare the verbs in study we will define what we will call *basic adicity* as a basic property of these verbs. We define basic adicity as the minimum number of arguments overtly required by a predicate to form well-formed, context independent sentences.

As stated in section 3.2.2 above, the incorporation of semantic components in Portuguese verbs of movement can result in the lexicalization of these components or in the incorporation of semantic restrictions on the semantic components they refer to. In the last case, and in comparison with the basic adicity of their hyperonym, this results in an increase of the number of overt arguments in hyponyms, whereas in the case of lexicalization it can result in the decrease of the number of overt arguments required by the hyponyms. Desirably, differences in the adicity of related verbs can be explained by the semantic properties of the verbs, directly related to lexicalization issues.

In what concerns Portuguese verbs of movement, the incorporation of semantic restrictions (+) on SOURCE and GOAL results in an increase of the number of arguments of the hyponyms, as exemplified in (29)a and b, whereas the lexicalization (++) of these components results in a decrease of the number of overt arguments of the hyponyms, in (30)a and b:

(29) a.

{mexer}v (change position); V(ARG$_1$, ARG$_2$);
Ex: O homem mexeu a cadeira
(The man moved the chair)

+GOAL

{pôr, colocar}v (*mexer* into a given position); V(ARG$_1$, ARG$_2$, **ARG$_3$**);
Ex: O homem pôs/colocou a cadeira **de/em pé**
(The man placed the chair up/in an upward position)

b.

{mover, deslocar}v (change location); V(ARG1, ARG2);
Ex: O homem moveu/deslocou a cadeira
(The man moved the chair)

+SOURCE

{tirar}v (*mover* from a given location); V(ARG1, ARG2, **ARG3**);
Ex: O homem tirou a cadeira **do armário**
(The man removed the chair from the closet)

(30) a.

{pôr, colocar}v (*mexer* into a given position); V(ARG1, ARG2, ARG3);
Ex: O homem pôs/colocou a cadeira de/em pé.
(The man placed the chair up/in an upward position)

++GOAL

{inclinar}v (*pôr* in an oblique position); V(ARG1, ARG2);
Ex: O homem inclicnou a cadeira (The man bent the chair)

b.

{tirar}v (*mover* from a given location); V(ARG1, ARG2, ARG3);
Ex: O homem tirou a cadeira do caixote.
(The man removed the chair from the box)

++SOURCE

{desencaixotar}v (*tirar* from a box); V(ARG1, ARG2);
Ex: O homem desencaixotou a cadeira (The man unboxed the chair)

The lexicalization of PATH, on the other hand, results in the increase of the number of arguments of hyponym verbs, usually corresponding to a new argument, which may be seen as an Obstacle[5] or GROUND referring argument:

(31) a.

{mover-se}v (move onelsef); V(ARG1);
Ex: O homem moveu-se. (The man moved himself)

++PATH

{circundar}v (*mover-se* around something); V(ARG1, **ARG2**);
Ex: O homem **circundou** a estátua.
(The man moved around the statue)

b.

{navegar}v *(mover-se* over water in a boat*);* V(ARG1);
Ex: O homem navegou. (The man navigated)

++PATH

{circum-navegar}v (*navegar* around something); V(ARG1, **ARG2**);
Ex: O homem circum-navegou **o continente**.
(The man circumnavigated the continent)

As defined for SOURCE and GOAL components, the incorporation of semantic restrictions on PATH also results in the requirement of a new argument referring the PATH of the movement, in Portuguese introduced by the preposition *por* (through), as the examples below show:

(32) a. {retroceder}$_V$ (move backwards, through the same path)

Ex.: O homem retrocedeu **pela estrada**. (the man moved backwards through the road)

[5] Obstacle arguments are here taken as defined in Fong & Fellbaum (2003). Briefly, Obstacle arguments refer to bounded areas; cannot be partitioned; enter (periphrastic) passive constructions; enter middle construction; and allow (-*ing*) nominalization.

b. {seguir}$_V$ (move oneself through a given path)
Ex.: O homem seguiu **pela estrada velha**. (the man moved through the old road.)

c. {circular}$_V$ (move oneself, cyclically, through a given path) (+MANNER)
Ex.: O sangue circula **pelas artérias**. (the blood circulates through the arteries)

d. {circular, transitar, andar}$_V$ (move oneself, usually, through a given path) (+MANNER).
Ex.: Os veículos circulavam/transitavam/andavam **pela estrada**.
(the vehicles circulated/moved around through the road)

The lexicalization or incorporation of semantic restrictions of other semantic components considered, namely, FIGURE, CAUSE, DIRECTION and INTENTION do not present typical patterns of change in the basic adicity of verbs. However, as previously demonstrated, all the semantic components considered are directly related to the argument structure of the verbs analyzed both in what concerns the predicates basic adicity and with regard to their selection properties.

3.2.4.3 Aktionsart properties

The Aktionsart properties (i.e. lexical aspectual properties of verbs) of verbs of movement can also be related to the incorporation of the semantic elements considered. In this section, we present our analysis of Portuguese verbs of movement in what concerns the lexicalization of the semantic elements considered and their correlation with Aktionsart shifts.

As it is commonly acknowledged, aspectual properties are greatly dependent on contextual factors (Moens 1987: 44), as the sentences in (33) demonstrate:

(33) a. John ran in the park (for a while). (*activity*)

b. John ran a mile (in less than four minutes). (*accomplishment*)

For this reason, in order to define the Aktionsart value of predicates we will also consider the basic adicity of the verbs in study, that is, the minimum number of arguments overtly required by a predicate to form well-formed, context independent sentences, as explained in the previous section.

We adopt the Vendler's (1967) four-class typology of – states, activities, accomplishments and achievements. In order to determine the basic Aktionsart value of our verbs, we followed Móia's (2000) set of criteria, taking into account both ontological and distributional properties, as schematized in Table 2, below.

Aktionsart class →		State	Activity	Accomplishment	Achievement
Ontological Properties	Temporal extendedness	non-punctual	non-punctual	non-punctual	punctual
	Homogeneity (subinterval property)	totally homogeneous	relatively homogeneous	heterogeneous	heterogeneous
	Nuclear Structure	one nuclear component	one nuclear component	preparatory process, culmination and consequent state	culmination (and possible consequent state)
Distributional Properties	Time adverbials (telicity)		**in* -adverbials *for* -adverbials	*in* -adverbials **for* -adverbials	**in* -adverbials **for* -adverbials
	Tense forms	**is V-ing*	*is V-ing*	*is V-ing*	
	Logical entailments		*is (now) V -ing → has V*	*is (now) V -ing → has not (yet) V*	

Table 2: Ontological and distributional properties of Aktionsart classes

The semantic components that are responsible for Aktionsart shifts of hyponym verbs in Portuguese verbs of movement are GOAL and SOURCE. The major lexicalization patterns found are described below, according to the Aktionsart classes established in Table 2.

(34) **V[activity]** $\xrightarrow{\text{+GOAL}}$ **V[accomplishment/achievement]**

a.{mover-se, deslocar-se}$_V$ [move oneself]

{voltar, regressar}$_V$[*mover-se*, again, to the start point]

b. O homem moveu-se/deslocou-se durante 10 minutos.
(The man moved for 10 minutes)

b′. O homem está (agora) a mover-se/deslocar-se → O homem moveu-se/deslocou-se
('The man is (now) moving → The man has moved)

c. O homem voltou/regressou em 10 minutos.
(The man returned in 10 minutes)

c′. O homem está (agora) a voltar/regressar → O homem (ainda) não voltou/regressou
(The man is (now) returning → The man has not (yet) returned)

(35) **V[activity]** **+SOURCE →** **V[accomplishment/achievement]**

a. {mover, deslocar}$_V$ [change location]

{tirar}$_V$ [*mover* from a given initial location]

b. O homem moveu/deslocou a caixa durante 10 minutos.
(The man moved the box for 10 minutes)

b'. O homem está (agora) a mover/deslocar a caixa → o homem moveu/deslocou a caixa.
(The man is (now) moving the box → the man has moved the box)

c. O homem tirou a caixa da rua em 10 minutos.
(The man took the box from the street in 10 minutes)

c'. O homem está (agora) a tirar a caixa da rua → o homem (ainda) não tirou a caixa da rua.
(The man is (now) taking the box from the street → the man has not (yet) taken the box from the street)

The lexicalization of SOURCE and GOAL, or the incorporation of semantic restrictions on these components, generally results in accomplishment type events. This seems to be the case given that the determination of a specific final location or position (GOAL) or initial location or position (SOURCE) establishes the final state of the event, shifting an activity into an accomplishment type event. However, in the set of verbs analyzed, there are also cases where the incorporation of SOURCE and GOAL results in achievement denoting verbs, as it is the case of the verbs *sair* (exit) and *entrar* (enter).

According to Moens (1987:45), shifts between activities and achievements can also occur. Intuitively, this seems possible in cases where the hyponym verb concerns a specific point of the activity denoted by the hyperonym that involves a consequent state, such as being in or out of some place, for instance. The concept denoted by the verb *sair* (exit) illustrates this case. *Sair* (exit) can be accurately described through the paraphrase "mover-se para fora de" (*move out of*), indicating a complex event structure involving a process (move) and a final state (being out of). Although not unproblematic, this intuition follows Pustejovsky (1995:160) proposal that achievement and accomplishment type events are composed of two subevents: a process and a final state. Accomplishments (such as *build*, for instance) have the process as head subevent, explaining the prominence of the process event, demonstrated by the entailment tests above and by the co-occurrence with *in*-adverbials. Achievements, on the other hand, have the final state as head subevent, which explains the impossibility its occurrence with *in*-adverbials, at least with the same reading as accomplishments, as we will demonstrate ahead.

This proposal is corroborated by Portuguese examples which show that verbs of movement such as *sair* (exit) and *entrar* (enter) can co-occur with *in*-adverbials with similar results to those of accomplishment denoting verbs such as *construir* (build):

(36) a. O ladrão construiu o esconderijo em 2 minutos. (≅ o ladrão demorou 2 minutos a construir o esconderijo)
(The thief built the hiding place in 2 minutes (≅ the thief took 2 minutes to build the hiding place))

b. O ladrão saiu do labirinto em 2 minutos. (≅ o ladrão demorou 2 minutos para sair do labirinto)
(The thief exited the maze in 2 minutes (≅ the thief took 2 minutes to exit the maze))

d. O ladrão entrou na caixa-forte em 2 minutos (≅ o ladrão demorou 2 minutos para entrar na caixa-forte)
(The thief entered the vault in 2 minutos (≅ the thief took 2 minutes to enter the vault))

In all the cases present above, the resulting readings concern the duration of the process – of actually building the hiding place, exiting the maze or entering the vault – that leads to the final state of the event.

The co-occurrence of these verbs with *for*-adverbials is also possible, although it results in a different reading from the one intended with the tests proposed in Table 2 above, in which *for*-adverbials test the atelicity of activities, or from the contrasts presented by Pustejovsky (1995) for accomplishments, in which *for*-adverbials refer to the duration of the preparatory process, in (37)a, and not to the duration of the final state, see (37)b and c. That is, achievement denoting verbs, occurring in sentences with *for*-adverbials, still denote an achievement type event, the duration final state being modified by the adverbial.

(37) a. O ladrão construiu o esconderijo durante 2 minutos. (≅ o ladrão esteve 2 minutos a construir o esconderijo/#O esconderijo esteve construído durante 2 minutos)
(The thief built the hiding place for 2 minutes (≅ the thief was building the hiding place for 2 minutes/#The hiding place was built for 2 minutes.)

b. O ladrão saiu do labirinto durante 2 minutos. (≅ o ladrão esteve fora do labirinto durante 2 minutos/#O ladrão esteve a sair do labirinto durante 2 minutos)
(The thief exited the maze for 2 minutes (≅ the thief was out of the maze for 2 minutes/#The thief was exiting the maze for 2 minutes))

c. O ladrão entrou na caixa-forte durante 2 minutos. (≅ o ladrão esteve dentro da caixa-forte durante 2 minutos/#O ladrão esteve a entrar na caixa-forte durante 2 minutos)

(The thief entered the vault for 2 minutes (≅ the thief was in the vault for 2 minutes/#The thief was entering the vault for 2 minutes))

The co-occurrence of the verbs *sair* (exit) and *entrar* (enter) with *for* and *in* adverbials is possible, although with different results from those achieved with accomplishment denoting verbs.

These data support the proposal which considers a complex internal structure of achievements, composed of two subevents – a process and a final state –, and accounts for the Aktionsart shift between activity denoting verbs and achievement denoting hyponyms. The determination of a specific final location or position (GOAL) or of an initial location or position (SOURCE) of the event sets a limit to the event, but also refers to a final state, thus explaining the Aktionsart shifts from activities to accomplishments or achievements.

Within the set of verbs analyzed, the lexicalization of the components SOURCE and GOAL, or the incorporation of semantic restrictions on these components, usually results in accomplishment type events, i.e. complex events in which the process subevent is prominent; however, in some cases, it results in achievement type events, i.e. complex events in which the final state subevent is prominent.

Although semantic lexicalization analysis can help to explain Aktionsart values shifting between hyperonym and hyponym verbs, it does not provide a direct account of all the phenomena observed. Aktionsart values can also be related to the argument structure of verbs: Aktionsart values shifts can, for instance, be induced by the expression of obstacle or GROUND arguments, as illustrated in (38) below.

(38) a. O homem navegou durante 10 dias/*em 10 dias.

(The man navigated for 10 days/in 10 days.)

a'. O homem está (agora) a navegar → o homem (já) navegou.

(The man is (now) navigating → the man has (already) navigated)

b. O homem circum-navegou a illha #durante 10 dias[6]/em 10 dias.

(The man circumnavigated the island *for 10 days/in 10 days)

b'. O homem está (agora) circum-navegar a ilha → o homem (ainda) não circum-navegou a ilha

(The man is (now) circumnavigating the island → the man has not (yet) circumnavigated the island)

[6] Acceptable only under an iterative reading.

Nonetheless, the regularity exhibited by the relation between lexicalization of SOURCE and GOAL and Aktionsart shifting, on the one hand, and the direct relation between semantic incorporation and differences in the number of arguments of the verbs observed, on the other, further motivate the need for richer informational structures at the lexical level, desirably contemplating argument and event structures that allow us to describe and explain the different semantic and syntactic properties of related verbs.

3.3 Conclusion

The analysis of the wordnet of Portuguese verbs of movement we built revealed lexicalization patterns different from those considered by Talmy (1985, 2000a) for Romance languages, corroborating Gutiérrez (2001) and Ibarretxe-Antuñano (2004) observations that Talmy's patterns do not conform to the data available. In fact, the study of Portuguese change of location and change of position verbs enabled us to propose a new set of semantic components introduced by the concepts denoted by these verbs, and tests to uniformly treat the data organized in our wordnet.

The organization of lexical items in a wordnet and the identification of the semantic components lexicalized allowed us to identify more accurately the specific semantic content of hyponyms with regard to their hyperonyms and its correlation to the semantic and syntactic behavior of verbs. The consequences are more or less visible, i.e. the lexicalization of semantic components affects inheritance to different degrees, namely in what concerns argument structure properties (number of arguments, argument subcategorization and semantic restrictions on the arguments selected), but also in what concerns Aktionsart properties, also being directly related to co-hyponyms compatibility.

The analysis of the data and the type of phenomena observed lead to the need to endow the WordNet model with richer informational structures. The hyponymy relation – that mediates between the lexical entries – assures the availability of lexical semantic structures, establishing that hyponyms denote the same concept as their hyperonyms, plus additional semantic contribution. However, and as "[hy]ponyms can be related to their superordinates along many semantic dimensions" (Fellbaum 1998:79), resulting in conceptual differences and properties that are not directly inherited by hyponym verbs, namely argument selection and event internal properties, as we have discussed in the last sections of this chapter, it is necessary to consider other levels of representation, contemplating argument and event structures, to accurately represent the verbs in study in the lexicon.

4. Modeling verbs in GL

4.0 Introduction

As discussed in chapter 3, verbs are involved in a large diversity of semantic and syntactic phenomena, whose treatment requires complex informational structures at the lexical entry level. The semantic and syntactic information stated in verbal lexical entries must, thus, account for the syntactic behavior of these lexical items, their polymorphic properties and sensibility to context. In our perspective, and having in mind the object of our research, the semantic and syntactic description of verbs at the lexical level should also be able to mirror and account for the differences between lexical-conceptually related verbs.

Given our goals, we chose to model the relevant linguistic information in the lexicon within the Generative Lexicon (GL) framework (Pustejovsky 1991, 1995). The levels of representation and the generative mechanisms considered in the GL model, on the one hand, and the range of phenomena considered in this work, on the other, motivate our choice. In addition, the organization of verbs of movement according to the lexical-conceptual relations established between them enables the sharing of relevant semantic and syntactic information within the net, resulting in a parsimonious modeling strategy. Moreover, as mentioned earlier, the integration of additional information in the Wordnet model does not compromise its global structure, rendering possible having a wordnet as the base for a generative lexicon.

Hence, in this chapter, we focus on the GL model, with particular attention to its representation levels, and discuss the elements to be considered regarding the modeling of the verbs in study, with the final goal of integrating GL structures in WordNet.PT.

4.1 The Generative Lexicon

The GL model, defined in Pustejovsky (1991, 1995), views the Lexicon as a complex system making up a crucial part of natural languages, and contemplating all the necessary information to account for the polymorphic properties of lexical items and their permeability to context. For this reason, lexical units are represented by informational structures, according to a finite set of rules, enabling the description of meaning in context and of the interface between lexical semantics and syntax.

The information is represented in a well-defined features structures system, in attribute-value matrixes (AVMs), whose values can be either atomic or, recursively, other AVMs. The systematic and declarative way in which the semantic content of lexical units is represented assures the coherent and recursive codification of the information, also enabling information sharing between structures, thus resulting in a relatively simple but adequate lexical modeling. Also, the operational devices defined in the model –generative mechanisms such as type coercion, co-composition and selective binding – are designed to account for the polymorphic properties of lexical items and their permeability to context, making use of the structured lexical information defined and represented in lexical entries.

4.1.1 Levels of representation in GL

The description of a lexical item in GL involves four levels of representation:

i) argument structure (A), which states the number and type of arguments of a lexical item;

ii) event structure (E), which refers to the properties of an event associated to a lexical item;

iii) qualia structure (Q), which provides the semantic objects that define the meaning of a lexical item;

iv) lexical inheritance structure (I), which lists the relations holding between a given lexical structure and other lexical structures in the lexicon.

A given lexical item, α, is thus represented by the information in these distinct levels, as described in (1).

(1) $\alpha = \langle A, E, Q, I \rangle$

4.1.1.1 Argument structure

Argument structure contemplates the definition of the semantic properties of the logical arguments of lexical items, as well as some syntactic mapping information, following the wide accepted conception of argument structure as highly structured (see Williams (1981), Grimshaw (1990), among others).

Besides referring to "traditional" arguments (i.e., independent variables with reference potential required by predicates to determine their own value), the notion of argument in GL also contemplates the semantic parameters that are part of the meaning of a lexical item. For instance, the lexical structure for a noun like *cat* has an argument (x) in its argument structure (ARGSTR) that corresponds to the object denoted by the lexical item, as formalized in the AVM in (2).

(2) cat

$$\begin{bmatrix} \text{ARGSTR} = \begin{bmatrix} \text{ARG}_1 = \text{x} : \text{cat} \end{bmatrix} \\ \ldots \end{bmatrix}$$

Pustejovsky (1995: 63-64) defines four types of arguments: true arguments; default arguments; shadow arguments; and true adjuncts.

True arguments (ARG_n) are "syntactically realized parameters of the lexical item" (Pustejovsky 1995:63), i.e., true arguments are necessarily expressed syntactically, their omission resulting in non grammatical sentences. For instance, the verb *construir* (build) has two true arguments in its argument structure (see (3)).

(3) a. [O João]$_{ARG1}$ construiu [uma casa]$_{ARG2}$.
([John]$_{ARG1}$ built [a house]$_{ARG2}$)

b. *Construiu uma casa.
(Built a house)

c. *O João construiu.
(John built.)

Default arguments (D-ARG_n) are "parameters which participate in the logical expressions in the qualia [structure], but which are not necessarily expressed syntactically" (Pustejovsky 1995:63), i.e. default arguments correspond to parameters used for the description of the semantic content of a given lexical item, but may or not be syntactically expressed. For instance, the concept denoted by the verb *build* requires the existence of construction material: wood, concrete, etc. And so, this parameter is considered a default argument, since, although required for the semantic description of the content of the lexical item, its syntactic realization is not required to guarantee the grammaticality of the sentence (see (4) below).

(4) a. O João construiu uma casa [de/em madeira]$_{ARG-D1}$.
(John built a house [of wood]$_{ARG-D1}$)

Shadow arguments (S-ARG$_n$) are, on their turn, "[p]arameters which are semantically incorporated into the lexical item. They can be expressed only by operations of subtyping or discourse specification" (Pustejovsky 1995:63-64) resulting, otherwise in semantically ill-formed sentences. *envenenar* (poison), kill with poison, in (5) and (6), illustrates this case:

(5) envenenar (poison)

$$\begin{bmatrix} \text{ARGSTR} = \begin{bmatrix} \text{ARG}1 = \text{x : entity} \\ \text{ARG}2 = \text{y : living being} \\ \mathbf{S\text{-}ARG1 = z : poison} \end{bmatrix} \\ \\ \text{EVENTSTR} = [\ldots] \\ \\ \text{QUALIA} = [\ldots] \\ \ldots \end{bmatrix}$$

(6) a. O João envenenou a Maria.
(John poisoned Mary)

b. #O João envenenou a Maria [com veneno]$_{\text{S-ARG1}}$.
(John poisoned Mary [with poison]$_{\text{S-ARG1}}$)

c. O João envenenou a Maria [com pesticida/arsénico]$_{\text{S-ARG1 subtypes}}$
(John poisoned Mary [with pesticide/arsenic]$_{\text{S-ARG1 subtypes}}$)

Finally, true adjuncts, the fourth type of argument considered in GL, are defined as the "parameters which modify the logical expression, but are part of the situational interpretation, and are not tied to any particular item's semantic representation" (Pustejovsky 1995:64). These arguments can be temporal or spatial adjuncts, for instance. True adjuncts are used in the characterization of classes of items, usually verb classes, associated to their general classification. For instance, individuated events can be modified by temporal or spatial adjuncts establishing a given point in time or space where the event occurred.

Given the goals of our work, namely the determination of the relevant semantic information to be considered in the lexical entries of Portuguese verbs of movement, true adjuncts will not be considered in the lexical representations of the verbs in study, as it is not relevant to state all the possible items that can co-occur with a given predicate, a set that, in theory, could amount to listing the entire lexicon.

The first three types of arguments considered – true arguments, default arguments and shadow arguments – account, thus, for the necessary semantic parameters required for the description of the semantic content of a given lexical item. However, and considering the syntactic mapping information attached to this representation level, as well as the syntactic criteria considered in the definition of argument types, we feel that some clarification on the definitions of arguments and argument types is in order.

The first clarification to be made concerns the definition of what constitutes an argument, i.e., which semantic objects are to be stated in the argument structure of a lexical entry. In a verb lexical structure, arguments correspond to semantic parameters, syntactically realized or not, that participate in the semantic contribution of the lexical item, but do not correspond to the semantic object denoted by it – the event –, in a clear contrast to what is considered to happen with nouns (see the example for *cat*, in (2) above). The semantic object corresponding to the denotation of a verb – always an event – is stated in another representation level: the event structure. In fact, the semantic contribution of event denoting lexical items, either verbal, nominal, adjectival, etc., is determined in the event structure, and not in the argument structure. This way, an argument could be defined as a non eventive semantic object that is part of the meaning of a lexical item. However, this rules out eventive arguments selected by some lexical items, such as verbs like *enjoy* or *begin*.

The syntactic mapping information level attached to the argument structure raises the second issue to be considered. Take, for instance, the argument structures of the noun *cat*, here in (7)a, and that of the noun *father*, a relational noun, in (7)b.

(7) a. cat

$$\begin{bmatrix} \text{ARGSTR} = \begin{bmatrix} \text{ARG1} = \text{x} : \text{cat} \end{bmatrix} \\ \ldots \end{bmatrix}$$

b. father

$$\begin{bmatrix} \text{ARGSTR} = \begin{bmatrix} \text{ARG1} = \text{x} : \text{human} \\ \text{D-ARG1} = \text{y} : \text{human} \end{bmatrix} \\ \ldots \end{bmatrix}$$

These representations correspond to a noun that does not select for any object, *cat*, and to a noun that may require the syntactic realization of its second argument, *father*. Considering similar representations for verbs, in (8) below, the syntactic mapping is considerably different: the syntactic realization of a verb with one true argument in its argument structure, as *run*, requires the realization of the verb plus one argument, typically occurring in subject position; the syntactic realization of a verb with two true arguments in its argument structure, like *eat*, requires the realization of the verb plus one argument in subject position and one argument in object position.

(8) a. run

$$\begin{bmatrix} \text{ARGSTR} = \begin{bmatrix} \text{ARG1} = \text{x} : \text{animal} \end{bmatrix} \\ \ldots \end{bmatrix}$$

b. eat

$$\begin{bmatrix} \text{ARGSTR} = \begin{bmatrix} \text{ARG1} = \text{x}:\text{animal} \\ \text{ARG2} = \text{y}:\text{physical object} \end{bmatrix} \\ \ldots \end{bmatrix}$$

Note, also, that event structure, as we will see ahead, does not contemplate the representation of the semantic object denoted by a verb, in the same way or with the same constraints on its syntactic realization as argument structure does.

Given that, on the one hand, argument structure is one of the representation levels that contemplates syntactic mapping, and that, on the other hand, it is also the representation level where it is possible to state the semantic objects that enter the description of the meaning of lexical items referring, namely, their semantic type, we proposed a new type of argument that accounts for the semantic object denoted by the lexical item itself. We will call these proper arguments[1] (P-ARG). This option directly results in a uniform representation of all POS, and has implications on the way lexical inheritance structure values are stated, as it will be discussed further ahead.

In the light of these considerations, we propose the following four types of argument to be considered in the argument structure of lexical items (leaving true adjuncts aside) and adding proper arguments to the list proposed by Pustejovsky (1995). Also, we introduce some slight reformulations in the definitions used.

i) Proper arguments (P-ARG_n): parameters of the lexical item semantic content that correspond to the semantic object denoted by the lexical item.

ii) True arguments (ARG_n): parameters of the lexical item semantic content whose syntactic omission can only be licensed when recoverable from context.

iii) Default arguments (D-ARG_n): parameters required by the logical expressions in the qualia structure of the lexical item that can be syntactically expressed by subtyping or specification processes.

iv) Shadow arguments (S-ARG_n): parameters incorporated in the lexical item that can only be syntactically expressed by subtyping or specification processes.

The definition of true arguments proposed here accounts for the semantic relevance of this type of arguments, i.e., they are an essential part of the meaning of lexical items, while licensing null argument contexts (see (9)), very common in Portuguese, but occurring also in English.

[1] We considered the designation of "proper argument" for two distinct sorts of reasons. Our first option, "self argument" would result in the same abbreviation as shadow arguments, S-ARG; "internal argument", on the other hand, could introduce additional noise since it corresponds to the syntactic notion of verbal internal objects, in opposition to external arguments, realized as subjects. For this reason, and following the analogy with the expression *proper noun*, we feel that the *proper argument* label is suitable.

(9) a. [e_i] Acordou cedo, [e_i] vestiu-se rapidamente e [e_i] saiu [e_j]. O porteiro do [hotel]$_j$ saudou-[o]$_i$ com um "bom-dia, [Sr. Mandela]$_i$."

[e_i] woke up early, [e_i] dressed quickly and [e_i] left [e_j]. the concierge of the [hotel]$_j$ greeted [him]$_i$ with a "good morning, [Mr. Mandela]$_I$"

'He woke up early, dressed quickly and left. The hotel concierge greeted him with a "good morning, Mr. Mandela".'

The slight reformulation of default arguments definition assures that the default arguments stated in the lexical entry of a given lexical item are those needed (required and not only that participate) to the representation of meaning content and not all the possible adjuncts that can co-occur with the lexical item at stake. Also, it states that the syntactic expression of this type of arguments is subject to similar constraints to those stated for shadow arguments: they can only be syntactically expressed through subtyping or specification processes, otherwise resulting in semantically ill-formed sentences. Consider, again, the case of the verb *build* that has a default argument in its argument structure referring to the construction materials required. The contrast between (10)a and (10)b shows that subtyping or specification is required to express default arguments syntactically.

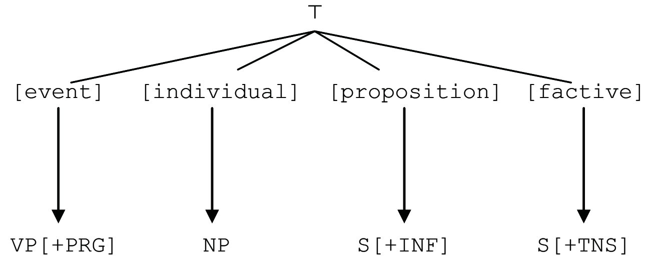
(10) a. #John built the house with materials.

b. John built the house with wood.

The reformulation of the types of argument considered in argument structure assures the coherent syntactic mapping of arguments for all POS.

The order by which arguments are stated within the argument structure also expresses constraints on the syntactic mapping. Typically, arguments are listed from the less oblique position (roughly corresponding to the subject position) to the more oblique one (corresponding, most of the times, to arguments expressed by prepositional phrases) (Pustejovsky 1995: 62-67). Based on the semantic type typology used for characterizing the arguments considered in the argument structure, Pustejovsky (1995: 136-140) takes the syntactic mapping of arguments a step further assuming that the top type of the hierarchy, $\top$, is subdivided into semantic types: event, individual, proposition and factive, corresponding to distinct canonical syntactic forms.

(11) Top nodes of GL semantic type lattice and respective canonical syntactic forms



(adapted from Pustejovsky (1995: 136))[2]

Using this type lattice, Pustejovsky (1995: 137) distinguishes the argument structure of verbs such as *like* and *enjoy*, here in (12)a and (13)a, respectively, accounting for the fact that the verb *like* can take all of ⊤ subtypes as arguments, whereas the verb *enjoy* takes only `event` type arguments (and its subtypes) as arguments (see (12)b and (13)b respectively).

(12) a. like

$$\begin{bmatrix} \text{ARGSTR} = \begin{bmatrix} \text{P-ARG}_1 = e_1 \\ \text{ARG}_1 = x:\text{ind} \\ \text{ARG}_2 = y:\text{T} \end{bmatrix} \\ \ldots \end{bmatrix}$$

b. She likes [watching movies]$_{event}$/[John]$_{individual}$/[John to watch movies with her]$_{proposition}$/[that John watches movies with her]$_{factive}$.

(13) a. enjoy

$$\begin{bmatrix} \text{ARGSTR} = \begin{bmatrix} \text{P-ARG}_1 = e_1 \\ \text{ARG}_1 = x:\text{ind} \\ \text{ARG}_2 = y:\text{event} \end{bmatrix} \\ \ldots \end{bmatrix}$$

b. She enjoys [watching movies]$_{event}$/[movies]$_{individual}$.

Besides stating the canonical syntactic forms of arguments, associating them to semantic types in the lattice, this typology also allows for multiple syntactic expressions of the arguments through type coercion, as we will see in section 4.1.2.3 ahead, and as it is exemplified in (13)b above. That is, *movies*$_{individual}$ in the sentence (13)b is coerced to `event` type, the type selected by the verb, being interpreted as *watching movies*.

[2] The type lattice and correspondent canonical syntactic forms presented in Pustejovsky (1995:136) leaves out the semantic type `property`, listed immediately before the graphic as a subtype of T. If we assume that this semantic type corresponds to the canonical syntactic form of Adjective phrases (see Pustejovsky 2001), the subsequent reasoning is weakened, since the verb *like* does not select an AdjP. This issue is not clarified by the author.

Note, however, that this representation still lacks some information on subcategorization properties, namely in what concerns subcategorization of prepositional phrases, as it is the case of SOURCE and GOAL denoting arguments of verbs of movement, for instance, or that of complements of adjective and adverbial phrases. Given that it is possible to have an AVM embedded in another AVM as value of a given feature, we propose to overtly state the lexical item that corresponds to, or introduces, the argument selected (in this case, the preposition), making use of the available structures in the lexicon, on the one hand, and accounting for the subcategorization properties of the lexical items, on the other. The AVM in (14) below illustrates two different cases: the verb *gostar* (like) that subcategorizes an argument introduced by the argument-marking preposition *de* (≅ of), selecting, as the English correspondent verb, for the top node $\top$, and the verb *sair* (exit) that selects a SOURCE denoting argument introduced by the preposition *de* (≅ from).

(14) a. gostar (*like*)

$$\begin{bmatrix} \text{ARGSTR} = \begin{bmatrix} \text{P-ARG}_1 = \text{e}_1 \\ \\ \text{ARG}_1 = \text{x : human} \\ \text{ARG}_2 = \begin{bmatrix} \text{de} \\ \text{ARGSTR} = [\text{ARG}_1 = \text{y : T}] \\ \ldots \end{bmatrix} \end{bmatrix} \\ \ldots \end{bmatrix}$$

c. sair (*exit*)

$$\begin{bmatrix} \text{ARGSTR} = \begin{bmatrix} \text{P-ARG}_1 = \text{e}_1 \\ \\ \text{ARG}_1 = \text{x : animated entity} \\ \text{ARG}_2 = \begin{bmatrix} \text{de} \\ \text{ARGSTR} = [\text{ARG}_1 = \text{y : location}] \\ \ldots \end{bmatrix} \end{bmatrix} \\ \ldots \end{bmatrix}$$

This way, we make use of the syntactic mapping information associated to the order of arguments within the argument structure (arguments are listed from the less oblique ones to the more oblique), assure a coherent mapping for nominal and verbal items (through the use of the P-ARG argument type), account for multiple syntactic forms of the arguments (through the use of the top semantic type $\top$), also specifying subcategorization constraints through the recursive use of available structures in the lexicon, represented by embedded AVM. This option requires, however, further analysis on the representation of prepositions, discussed in detail in the next chapter.

4.1.1.2 Event structure

The study of verbal predicates, as demonstrated for instance in section 3.2.4.3 of the previous chapter, shows that the determination of the internal structure of predicates – the event structure – is necessary to account for divergent semantic and syntactic behaviors: "an event structure can provide a distinct and useful level of representation for linguistic analysis involving the aspectual properties of verbs, adverbial scope, the role of argument structure, and the mapping from lexicon to syntax" (Pustejovsky 1991:47).

Contrary to the atomic view on event structure (Davidson 1967), Pustejovsky (1991, 1995) assumes that events can be non-atomic, i.e., events can be constituted by other subevents. The event structure, thus, consists in the representation of the internal structure of a given predicate, including the list of events associated to it, partial order constraints, overlapping and inclusion of subevents, and the definition of the event head.

The event structure (EVENTSTR) is constituted by the list of event arguments (E_1, ..E_n) that correspond to the subevents of a given event (in our proposal, represented as the proper argument), by the attribute Restrictions (RESTR) whose value represent the temporal order constrains holding between subevent arguments, and by the attribute Head (HEAD) which allows the representation of foregrounding and backgrounding of event arguments (see Pustejovsky (1995: 72)).

(15)

$$\alpha \\ \begin{bmatrix} \text{EVENTSTR} = \begin{bmatrix} \text{E}_1 = \text{e}_1 : \text{process} \\ \text{E}_2 = \text{e}_2 : \text{state} \\ \text{RESTR} =< \alpha \\ \text{HEAD} = \text{e}_2 \end{bmatrix} \\ \ldots \end{bmatrix}$$

Event arguments, as can be seen in (15) above, can have different types: states, processes and transitions. According to Marrafa (1993:27-28), these three types of events are defined as follows:

i) a state is an atomic event, non evaluated regarding any other, i.e., given a time interval I, a state is an event that is verified in all I;

ii) a process is a sequence of identical events, i.e., given a time interval I, a process is a sequence of events that is verified in all the intervals of I;

iii) a transition is an event evaluated regarding another event, i.e., in a time interval I, a transition is composed of a process and of a final state, which is different from the initial one.

Transitions are binary complex events. That is, they are composed of a preparatory process that leads to a final state, different from the initial one, that, although being entailed, is not part of the transition (Marrafa 1993). The sentences in (16) illustrate this case:

(16) a. John healed the dying cat/#healthy cat.

b. John cleaned the clean house.

c. #John built the built/ruined house.

In (16)a, the occurrence of the verb *heal* with the modifier *healthy* is not allowed since it specifies an initial state that is equal to the final state of the event denoted by the verb. In a first observation, this could indicate that the initial state is also a subevent of the transition. However, the same is not true for the verb *clean*, in (16)b, where *clean* can also specify the initial state of the object, also seemingly equal to the final state of the event. The contrast here seems to result from the creation notion of some transition denoting verbs, i.e., from the fact that the transitions denoted by some verbs refer to a final state of existence of the object, entailing as initial state the non-existence of the object. Verbs that denote transitions that do no refer to a final state of existence of the object, such as *clean*, do not show this restriction. Verbs such as *build*, in (16)c, on the contrary, do not allow the specification of the initial state of the object, since it is semantically odd to characterize a non-existent object. Verbs such as *heal* fall somewhere in between, i.e., the event denoted by the verb entails a non existing initial state of non-health and result in a final state of health, but it is possible to characterize the object since it exits and does not result from the event. However, the specification of this initial state of the object must be semantically coherent with the initial state entailed by the verb (see Marrafa (1993) for a detailed discussion on the subject).

Typically, the order by which event arguments are listed in the event structure reflects sequential order restrictions: the first event argument listed corresponds to the first subevent of the complex event, and so on. However, and to account for other relations holding between subevents, the values of Restrictions can describe part of relation ($\preccurlyeq$); strict partial order ($<$); overlapping ($\circ$) and inclusion ($\sqsubseteq$). This way, Pustejosvky (1995: 69-71) defines three temporal ordering relations realized in natural languages.

i) **Exhaustive ordered part of** relation, $<_{\alpha}$, in which e_3 is a complex event constituted by e_1 and e_2 subevents, logical parts of e_3, e_1 precedes e_2, and e_1 and e_2 are the only subevents of e_3, as stated in the following definition:

(17) a. $[_{e3}\ e_1 <_{\alpha} e_2] =_{def} <_{\alpha} (\{e_1, e_2\}, e_3)$

b. $\forall e_1, e_2, e_3\ [<_{\alpha} (\{e_1, e_2\}, e_3\}) \leftrightarrow e_1 \preccurlyeq e_3 \wedge e_2 \preccurlyeq e_3 \wedge e_1 < e_2 \wedge$
$\forall e[e \preccurlyeq e_3 \rightarrow e = e_1 \vee e_2]]$

The verb *build* denotes an event with such a structure: it is composed of two subevents, one preceding the other.

ii) **Exhaustive overlap part of** relation, $\circ_\alpha$, in which e_3 is a complex event constituted by e_1 and e_2 subevents, logical parts of e_3, e_1 temporally includes e_2, e_2 temporally includes e_1, both e_1 and e_2 being temporally included in e_3, and e_1 and e_2 being the only subevents of e_3, as stated in the definition in (18):

(18) a. $[_{e3}\ e_1 \circ_\alpha\ e_2] =_{def}\ \circ_\alpha (\{e_1, e_2\}, e_3)$

b. $\forall e_1, e_2, e_3\ [\circ_\alpha (\{e_1, e_2\}, e_3\}) \leftrightarrow e_1 \preccurlyeq e_3 \wedge e_2 \preccurlyeq e_3 \wedge e_1 \sqsubseteq e_2 \wedge e_2 \sqsubseteq e_1 \wedge \exists e[e \sqsubseteq e_1 \wedge e \sqsubseteq e_2 \wedge e = e_3] \wedge \forall e[e \preccurlyeq e_3 \rightarrow e = e_1 \vee e_2]]$

The verb *drive* denotes an event with such a structure: it involves two subevents (driving and motion) that occur simultaneously.

iii) **Exhaustive ordered overlap** relation, $<\circ_\alpha$, in which e_3 is a complex event constituted by e_1 and e_2 subevents, logical parts of e_3, e_1 temporally includes e_2, the initial part of e_1 precedes the initial part of e_2, the ending parts of e_1 and e_2 being simultaneous, and e_1 and e_2 being the only subevents of e_3, as stated in (19) below:

(19) a. $[_{e3}\ e_1 <\circ_\alpha\ e_2] =_{def}\ <\circ_\alpha (\{e_1, e_2\}, e_3)$

b. $\forall e_1, e_2, e_3\ [<\circ_\alpha (\{e_1, e_2\}, e_3\}) \leftrightarrow e_1 \preccurlyeq e_3 \wedge e_2 \preccurlyeq e_3 \wedge e_1 \circ e_2 \wedge \text{init}(e_1) < \text{init}(e_2) \wedge \text{end}(e_1) = \text{end}(e_2) \wedge \forall e[e \preccurlyeq e_3 \rightarrow e = e_1 \vee e_2]]$

A verb denoting a complex event in which the second event argument is properly included in the first event argument can illustrate this structural relation. Pustejovsky (1995: 71) exemplifies this relation with the event structure of the verb *walk*, the subevent of moving the leg corresponding to e_2, and moving the hole body corresponding to e_1, resulting in the process of walking. However, we feel that this fine-grained characterization of processes such as *walk*, given our definition of a process as a sequence of identical events, is not sufficiently motivated to be considered necessary for the semantic representation of this type of event structure, i.e. the event structure presented corresponds to the repeated event that constitutes a process. As we discuss in the chapter 7 of this dissertation, the definition of the subevents of process type events seems only relevant when these correspond to conceptually individuated events on their own, and are not necessarily reflected in the event structure of lexical items.

Consider, for instance, the event denoted by the verb *dactilografar* (≅ copy using a typewriter, type). This event can be described as a complex process composed of an initial process of reading/hearing, followed by the mental processing of the information read/heard simultaneously occurring with the process of pressing keys in a typewriter. The verb *dactilografar* (≅ type) can, thus, be represented as having the following event structure:

(20) dactilografar(≅ type)

$$\left[\text{EVENTSTR} = \begin{bmatrix} \text{E}_1 = \text{e}_1 : \text{process} \\ \text{E}_2 = \text{e}_2 : \text{process} \\ \text{E}_3 = \text{e}_3 : \text{process} \\ \text{RESTR} = \text{e}_1 < \text{e}_2, \text{e}_2 \circ \text{e}_3 \end{bmatrix}\right]$$

However, besides contradicting Pustejovsky (1995) assumption that complex events are necessarily composed of a finite list of two subevents, this formulation is also contrary to the example given by the author for the relation of exhaustive ordered overlap linking the constituent subevents of a process, since these subevents constitute the event whose repetition forms the process of typing. Otherwise, the event denoted by *type* would not be a process.

As result of the discussion presented above, the base event structures used in this work to represent the event types considered, namely states, processes and transitions, are respectively formalized in (21) below:

(21) a. state

$$\left[\text{EVENTSTR} = \left[\text{E}_1 = \text{e}_1 : \text{state}\right]\right]$$

b. process

$$\left[\text{EVENTSTR} = \left[\text{E}_1 = \text{e}_1 : \text{process}\right]\right]$$

c. transition

$$\left[\text{EVENTSTR} = \begin{bmatrix} \text{E}_1 = \text{e}_2 : \text{process} \\ \text{E}_2 = \text{e}_3 : \text{state} \\ \text{RESTR} = <\alpha \end{bmatrix}\right]$$

As atomic events, states do not have a complex event structure. Although being complex events, processes are basically represented as atomic since the sequence of identical events that form them is not directly decomposable. In this perspective, we propose the same event structure for processes, such as *run*, whose subevents (move one leg, projecting the body forth, moving the other leg, projecting the body forth, move one leg, etc.) cannot be conceptually individuated, and for complex processes, such as *breath*e, for instance, that has inhale and exhale, both conceptually individuated, as subevents.

Besides event arguments and structural order constraints, the event structure level of representation allows for the definition of the head event of a complex event structure. This attribute, Head, enables the statement of the prominence of the subevents in the event structure, which, besides allowing accounting for different Aktionsart values of the predicates, is used by Pustejovsky (1995) to account for the relation between the logical senses of causative/inchoative predicates and that of related verbs such as *buy/sell*, *give/take*, etc., for instance.

Given the event structures considered in (21) above, the definition of the head of the event reflects the Aktionsart properties of accomplishments and achievements, both transition type events: accomplishment type events are transitions in which the process subevent is prominent, see (22)a, whereas achievements type events are transitions in which the final state is the prominent subevent, see(22)b.

(22) a. accomplishment

$$\left[\text{EVENTSTR} = \begin{bmatrix} \text{E1} = \text{e1} : \text{process} \\ \text{E2} = \text{e2} : \text{state} \\ \text{RESTR} = < \alpha \\ \textbf{HEAD} = \textbf{e1} \end{bmatrix}\right]$$

b. achievement

$$\left[\text{EVENTSTR} = \begin{bmatrix} \text{E1} = \text{e1} : \text{process} \\ \text{E2} = \text{e2} : \text{state} \\ \text{RESTR} = < \alpha \\ \textbf{HEAD} = \textbf{e2} \end{bmatrix}\right]$$

This configuration explains the different readings resulting from the modification by *for*-adverbials of accomplishment and achievement denoting predicates. In the first case, see (23)a, the subevent modified is the process, the head event, whereas in the second case, as illustrated in (23)b, the subevent modified is the final state, also the head event in this case.

(23) a. John climbed the Everest <u>for an hour</u>.

b. My terminal died <u>for two days</u>. (examples from Pustejovsky (1995: 74))

The definition of the head event can also encode "mirror" events such as the ones denoted by *buy* and *sell*. These two verbs refer to the same event, which has at least two subevents: one where one person gives something to another person in exchange for money and other where one person acquires something in exchange for money, informally, EVENTSTR= [E_1=e_1: acquire, E_2 = e_2: give, RESTR= $\circ_\alpha$]. Depending on the focus on these subevents, we have *buy* (EVENTSTR= [E_1=e_1: acquire, E_2 = e_2: give, RESTR= $\circ_\alpha$, **HEAD = $\mathbf{e_1}$**]) and *sell* (EVENTSTR= [E_1=e_1: acquire, E_2 = e_2: give, RESTR= $\circ_\alpha$, **HEAD = $\mathbf{e_2}$**]).

Pustejovsky (1995) also relates event headedness with the treatment of verbal polysemy, namely in what concerns causative/inchoative readings. Polysemy occurs when the lexical expression of a verb is not specified regarding the head event in the event structure. Consider, for instance, the AVM proposed for the verb *sink*, here in (24)a. Not being specified in the lexicon, this representation allows for the contextual determination of the head event, resulting in one case in a causative verb, the process subevent being the head event of the structure (see (24)b) and, in the other, in a inchoative verb, when the final state is the head event of the structure (see (24)c).

(24) a. sink

$$\begin{bmatrix} \text{EVENTSTR} = \begin{bmatrix} \text{E1} = \text{e1}: \text{process} \\ \text{E2} = \text{e2}: \text{state} \\ \text{RESTR} = <\alpha \\ \text{HEAD} = \end{bmatrix} \\ \ldots \end{bmatrix}$$

b. The navy sunk the boat. (EVENTSTR=[..., HEAD = $\mathbf{e_1}$])

c. The boat sunk. (EVENTSTR = [..., HEAD = $\mathbf{e_2}$])

Marrafa & Moura (2005) notice, however, that syntactic alternations may be related to other semantic properties besides event headedness. The examples in (25)a, b and c demonstrate that the alternative syntactic constructions in which Telic causative verbs such as *preparar* (≅ prepare), *planear* (≅ plan) or *organizar* (≅ organize) can occur emerge from the specification of a third argument (see (25) b′ and c′).

(25) a. O treinador preparou o atleta para a competição.
(The coach prepared the athlete for the competition.)

a′. #O treinador preparou a competição.
(The coach prepared the competition)

b. O João planeou todos os detalhes para a festa.
(João planned all the details for the party)

b′. O João planeou a festa.
(João planned the party)

c. O João organizou tudo para a festa.
(João organized everything for the party)

c′. O João organizou a festa.
(João organized the party)

(Examples adapted from Marrafa & Moura 2003: 55, 57)

The sentences in (25)a′, b′ and c′ show that it is only possible to have alternative syntactic realizations of arguments, in this case the promotion of the third argument (*the competition,* and *the party,* in the examples above) to the object position, if the second argument of the verb is not semantically strong, as it is the case of (25)a′, which results in a different sentence altogether. Hence, this alternation does not emerge from the event head specification but rather from the semantics of the arguments.

Finally, the event structure level of representation is not necessarily bound to verbal semantic representations. Event denoting nouns such as *war*, *construction*, etc., or adjectives for instance, necessarily require this level of representation.

4.1.1.3 Qualia structure

The qualia structure (QUALIA) is the level of representation in which the semantic content of lexical items is explained, i.e. where the set of properties or events associated to the meaning of a given lexical item is stated: "What qualia structure tells us about a concept is the set of semantic constraints by which we understand a word when embedded within the language. The *mode* of explanation that characterizes a word as denoting a particular concept is potentially distinct from the manner in which that word is used in language." (Pustejovsky 1995: 86). Qualia structure is, thus, the level of representation that concerns the expression of semantic relational structures, accounting for the polymorphic behavior of lexical items and, simultaneously, for their context sensibility, i.e. the creation of new meanings in context.

This representation level is composed of four attributes, the qualia roles, and their values, whose determination is not compulsory: for a given lexical item, only the relevant values of pertinent qualia roles must be determined. Based on Aristotle's modes of explanation, Pustejovsky (1995) defines the four qualia roles as follows:

i) Constitutive role (CONST): attribute whose values express the relation between a given object and its constituents or parts, such as material, weight or proper parts;

ii) Formal role (FORMAL): attribute whose values establish what distinguishes a given object within its semantic domain, typically its basic category description: the sortal typing of the proper-argument or the relation between proper-arguments of different types;

iii) Telic role (TELIC): attribute whose values concern the information about the function or purpose of the object, such as the intention of an agent in performing a given action or the built-in function of a given object;

iv) Agentive role (AGENTIVE): attribute whose values determine the origin of the object, such as its creator, origin (artifact or natural kind) or causal chain.

The values of qualia roles are stated through logical expressions, with well-defined types and relational structures, indicating the proper binding of the predicating term. So, the noun *novel* for example, will have the qualia structure exemplified in (26) below.

(26) a. novel

$$\begin{bmatrix} \text{ARGSTR} = [\text{P-ARG}_1 = \text{x}] \\ \\ \text{QUALIA} = \begin{bmatrix} \text{CONST} = \text{has\_part}(\text{x}, \text{y}:\text{narrative}), \ldots \\ \text{FORMAL} = \text{book}(\text{x}) \\ \text{TELIC} = \text{read}(\text{e}, \text{z}:\text{human}, \text{x}) \\ \text{AGENTIVE} = \text{write}(\text{e}, \text{w}:\text{writer}, \text{x}) \end{bmatrix} \\ \ldots \end{bmatrix}$$

This formulation, besides capturing the semantic and conceptual relations associated to the meaning of the lexical item, accounts for the interpretations licensed by local semantic and syntactic contexts as well as modification of qualia values and type coercion phenomena. Consider, for instance, the sentences in (27)a and b and their readings.

(27) a. This is a good novel. (≅ this is a novel that reads well, well written, with a good narrative)

b. Mary began the novel. (≅ Mary began reading/writing the novel)

In (27)a, it is possible to see how the adjective modification refers to the information stated in the qualia structure of the noun *novel*, exemplifying the occurring modification process. *Good* modifies the several values of the qualia roles of the noun *novel*.

The sentence in (27)b, on its turn, shows how the information stated in the qualia structure enables type coercion as well as the computation of the meaning of the sentence. In this case, the verb *begin* selects an argument of the type event. However, it can occur with nouns such as *novel*, by coercing it to an event type argument (from `individual` to `event`), recovering the events stated in its qualia structure.

Qualia structure also allows predicting how co-composition processes take place, by establishing which values are considered in a underspecified representation. Take, for instance, the verb *build*. As stated in section 4.1.1.1 above, *build* includes a default argument that refers to the material used in the process of building. However, this material depends on the object created: if it is a house it can be wood, concrete, etc., if it is a car it is metal, rubber, etc. The AVM in (28) (adapted from Pustejovsky (1995:82)) shows how the three levels of representation interact and how qualia values can be accurately used for meaning computation using underspecified representations. The default argument of the verb, D-ARG$_1$, is related to the logical object, ARG$_2$, through the unification of the values expressed in the Constitutive role of ARG$_2$, represented in the AVM by the numerical index $\underline{3}$.

(28) build

$$\begin{bmatrix} \text{ARGSTR} = \begin{bmatrix} \text{P - ARG}_1 = e_1 \\ \text{ARG}_1 = \boxed{1}\begin{bmatrix} \ldots \\ \text{FORMAL} = \text{animated\_entity}(x) \end{bmatrix} \\ \text{ARG}_2 = \boxed{2}\begin{bmatrix} \ldots \\ \text{CONST} = \boxed{3} \\ \text{FORMAL} = \text{physical\_object}(y) \end{bmatrix} \\ \text{D - ARG}_1 = \boxed{3}\begin{bmatrix} \ldots \\ \text{FORMAL} = \text{material}(z) \end{bmatrix} \end{bmatrix} \\ \text{EVENTSTR} = \begin{bmatrix} E_1 = e_1 : \text{process} \\ E_2 = e_2 : \text{state} \\ \text{RESTR} = <\alpha \\ \text{HEAD} = e_1 \end{bmatrix} \\ \text{QUALIA} = \begin{bmatrix} \text{FORMAL} = \text{exist}(e_2, \boxed{2}) \\ \text{AGENTIVE} = \text{build\_act}(e_1, \boxed{1}, \boxed{3}) \end{bmatrix} \\ \ldots \end{bmatrix}$$

As shown in the example above, the qualia structure for verbal expressions does not conform exactly to the definitions proposed for qualia roles by Pustejovsky (1995). This seems to be related to the fact that eventive concepts are not adequately described by ontological categories, as are non-eventive ones. For this reason, we will discuss verbal qualia structure in the next subsection.

4.1.1.3.1 Verbal qualia structure

The first issue to consider regarding the qualia structure of verbs is related to the values of the qualia roles that are licensed for these items. Verbal qualia structure is fulfilled with the semantic predicates that establish the relations between event arguments (stated in the event structure) and the arguments (defined in the argument structure) of a given verb. Pustejovsky (1995) uses only two of the four qualia roles in the representation of verbal semantics: the Formal role and the Agentive role.

The Formal role is used to establish the semantic predicates of states and stative processes (i.e., processes that do not require an agentive argument, such as *sleep*), since they can be seen as corresponding to an existing state of affairs, without reference to how the event comes about (see Pustejovsky (1995:79)). Also, it is in the Formal role that the author encodes the semantic predicate corresponding to the final state in transitions.

The Agentive role, related to the causal chain or origin of the event, is used to state the preparatory process that leads to the final state in transition denoting verbs, and the processes in active process denoting verbs such as *run*.

This distinction between active and passive processes and between the qualia roles used for their codification explains, according to the author, why it is possible to explicitly specify the resulting state (as value of the Formal role) for active processes such as *run* (see (29)), whereas the same is not possible for passive processes, such as *sleep*, which are limited to modification by durative adverbials: the Formal role value is already filled-in.

(29) run home/ run to the store

$$\begin{bmatrix} \ldots \\ \text{QUALIA} = \begin{bmatrix} \text{FORMAL} = \text{at\_home}(e_2, \boxed{1}) \ / \text{at\_the\_store}(e_2, \boxed{1}) \\ \text{AGENTIVE} = \text{run\_act}(e_1, \boxed{1}) \end{bmatrix} \\ \ldots \end{bmatrix}$$

Note, however, that the arguments presented for considering different qualia roles for active and passive processes and the inferred assumption that verbal qualia roles can only have one value are questioned by at least two examples used in Pustejovsky (1995), namely examples with verbs *drive* and *float* ((Pustejovsky 1995:114, 125-126, respectively).

The first case, reproduced here in (30)a, is represented as a complex process with two simultaneous process subevents, *move* and *drive*, values of the Formal and Agentive qualia roles. Assuming that for passive processes it is not possible to specify the final state since the Formal role value is already filled-in, as opposed to what occurs with active processes, sentences such as the ones in (30)b below should be ruled out, given that the Formal role is also already filled-in for the verb *drive*. However, this is not the case.

(30) a. drive

$$\begin{bmatrix} \text{ARGSTR} = \begin{bmatrix} \text{P-ARG}_1 = e_1 \\ \text{ARG}_1 = x : \text{human} \\ \text{ARG}_2 = y : \text{vehicle} \end{bmatrix} \\ \\ \text{EVENTSTR} = \begin{bmatrix} E_1 = e_2 : \text{process} \\ E_2 = e_3 : \text{process} \\ \text{RESTR} = \circ\alpha \end{bmatrix} \\ \\ \text{QUALIA} = \begin{bmatrix} \text{FORMAL} = \text{move}(e_3, x) \\ \text{AGENTIVE} = \text{drive\_act}(e_2, x, y) \end{bmatrix} \\ \end{bmatrix}$$

b. John drove home/to the store.

The case of the verb *float* also reveals an inconsistency of the proposal, but of the opposite type. Although there is no clear indication of what constitutes an active *versus* a passive process, it seems to be assumed that an active process requires a dynamic agent whose voluntary action is necessary for the occurrence of the process – hence the statement of the correspondent semantic predicate in the Agentive role –, whereas a passive process does not require any agentive action to occur – being thus considered as a current state of affairs with no reference to its origin –, such as in the case of the process denoted by the verb *sleep*. In the light of these considerations, the event denoted by the verb *float*, not requiring an active agent, should be stated as a value of the Formal role (note that the author characterizes the event denoted by the verb *float* as a state (Pustejovsky 1995: 125) thus being a stative property). And yet, in order to conform to the data, namely the possibility of specifying a final state as the value of the Formal role (see (31)), the verb *float* is represented as having the denoted process as the value of the Agentive role (see Pustejovsky 1995: 125-126).

(31) The bottle floated to the shore/into the sewer.

The distinction between what may constitute the value of the Formal and Agentive roles is consistent with the previous definition of these attributes – Formal: attribute whose values establish what distinguishes a given object within its semantic domain, typically its basic category description; and Agentive: attribute whose values determine the origin of a given object, such as its creator, origin (artifact or natural kind) or the causal chain –, although some slight adjustments may contribute to a more general definition that better suites all POS.

However, we consider that the assumption that verbal qualia roles can only be filled in with one semantic predicate is not sufficiently motivated, given that it rules out well-formed sentences such as the one presented in (30)b above[3].

Although not considered in Pustejosvky (1995), the Telic role of verbal qualia structures can also be used to describe events that correspond to the specific function of the event denoted by some verbs (Marrafa & Moura 2003). Marrafa & Moura (2003) explain that this is case of Telic Causative verbs such as *preparar* (≅ prepare) and propose that the function event denoted by these verbs is stated as the value of the Telic role. The verb *preparar* (≅ prepare) denotes a complex event that has an active process, described as the value of the Agentive role – the activity of preparing something–, a resulting state, described in the Formal role – the result of the preparation activity-, and a function process, described as the Telic role value – the process for which the preparation activity is realized – as subevents, represented here in (32) (AVM from Marrafa & Moura (2003: 557)).

[3] We are following the reasoning proposed by Pustejosky (1995), in the sense that the examples presented refer to phrases and not lexical entries. This way, the author intends to show how compositionally works. However, the semantic predicates introduced by goal denoting phrases are necessarily not considered in the lexical entry of the verb *run*.

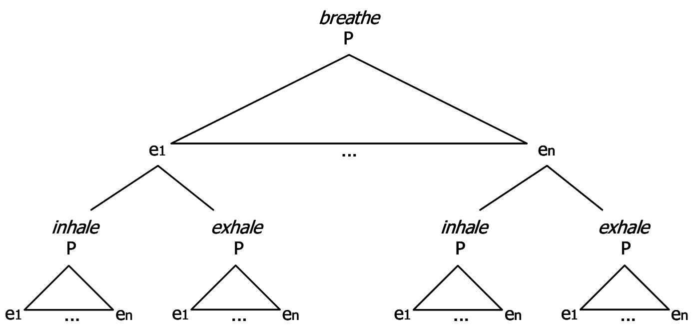
(32) preparar

$$\left[\begin{array}{ll}
\text{ARGSTR} = & \left[\begin{array}{l} \text{ARG1} = \boxed{1}[\text{animate_ind}] \\ \text{ARG2} = \boxed{2}[\text{entity}] \\ \text{ARG2} = \boxed{3}[\text{event}] \end{array}\right] \\
 & \\
\text{EVENTSTR} = & \left[\begin{array}{l} \text{E1} = \text{e}_1 : \text{process} \\ \text{E2} = \text{e}_2 : \text{state} \\ \text{E3} = \text{e}_3 : \text{process} \\ \text{RESTR} = \text{e}_1 < \text{e}_2 ; \text{e}_1 < \text{e}_3 \\ \text{HEAD} : \text{e}_1 \end{array}\right] \\
 & \\
\text{QUALIA} = & \left[\begin{array}{l} \text{telic_causative_lcp} \\ \text{FORMAL} = \text{prepare_result}(\text{e}_2, \boxed{2}) \\ \text{AGENTIVE} = \text{prepare_act}(\text{e}_1, \boxed{1}, \boxed{2}) \\ \text{TELIC} = \text{prepare_for}(\text{e}_1, \text{e}_3, \boxed{1}, \boxed{2}, \boxed{3}) \end{array}\right]
\end{array}\right]$$

Given the former definitions of the Formal and Agentive roles for verbal representations and the general definition of the Telic role, this is the coherent option. Thus, Telic role is also relevant in verbal qualia structures.

Finally, the Constitutive role seems to be overlooked since the event structure already establishes the subevents that compose the denoted event. Let us consider, for instance, the verb *breathe* that denotes a process type event, which can be described as having the conceptually individuated processes of *inhaling* and *exhaling* as subevents. These subprocesses, in fact, do not express a current state of affairs but rather the events that constitute the event whose repetition originates the process denoted by *breathe*, as depicted in (33) below.

(33)



As explained before, these events are not direct subevents of the process denoted by *breathe* and thus are not expressed in the event structure, being, nonetheless, related to the concept denoted by this verb as its parts. This way, these subevents may constitute values of the Constitutive role (attribute whose values express the relation between a given object and its constituents or parts). Also, it seems more accurate to state them as values of the Constitutive role than as values of the Formal role.

This way, it is possible to consider that the distribution of semantic predicates in verbal qualia structures is defined according to the facets described by qualia roles that, when applied to verbs, refer to:

i) Formal role values: semantic predicates corresponding to actual states of affairs, typically states;

ii) Agentive role values: semantic predicates involved in the causal chain of complex events, typically active processes;

iii) Telic role values: semantic predicates that express the specific function of the semantic object (see Marrafa & Moura 2003)

iv) Constitutive role values: semantic predicates that constitute the event whose repetition forms a process.

In sum, it seems to us that the simple adjustment of the definitions of qualia roles, in order to also comprise properties of events, is a step further for a more coherent and consistent treatment of all POS, enhancing the model.

Verbal qualia structure, given the nature of verbs and specifically the property of headedness of subevents, has yet another significant difference with regard to nominal qualia structure. Specifically, the event head definition in a complex event allows the foreground of a given subevent and of its respective specification in the qualia structure, mapping the semantic expression of the predicate of the head event to the appropriate syntactic form. Pustejovsky (1995) establishes that the value of the qualia role corresponding to the head event is projected to the s-structure, specifying the grammatical functions appropriate to the arguments of a given verb:

(34) Projection of Qualia roles values in s-structure

$V(x,y) \rightarrow x{:}_{\text{SUBJ}},\ y{:}_{\text{OBJ}}$

"If one normally thinks of projection as specifying the appropriate grammatical functions to the arguments of a lexical item, such as [(34)] then it is clear what the task for a qualia-based representation is; namely to project from multiple semantic expressions to the appropriate grammatical functions in syntax" (Pustejovsky 1995:101).

Assuming this mapping function, the author defines two possible projections from a given qualia role (Q_α), considering two-place predicates (in (35)a.) and one-place predicates (in (35)b). Note that these projections consider only non oblique positions in syntax, i.e. subject and object positions.

(35) a. Q_i: R(e_1, x, y) $\rightarrow$ x:SUBJ, y:OBJ

b. Q_j: P(e_2, y) $\rightarrow$ y:SUBJ

The determination of the head event of a complex event will, thus, define the projectable qualia role, acting as a filter: the head event, *e**, projects the template associated with its semantic predicate established as the value of a given role in the qualia structure. For instance, a verb such as *sink* can have the two following mappings:

(36) a. $Q_{Agentive}$:sink_act(e_1*, x, y) $\rightarrow$ x:SUBJ, y:OBJ

b. Q_{Formal}: final_state(e_2, y) $\rightarrow$ *shadowed*

c. The navy sunk the boat.

(37) a. Q_{Formal}: final_state(e_2*, y) $\rightarrow$ y: SUBJ

b. $Q_{Agentive}$:sink_act(e_1, x, y) $\rightarrow$ *shadowed*

c. The boat sunk.

To assure the correct mapping from the qualia structure of a lexical item, the values of the qualia structure must be *saturated* at syntax, i.e. the arguments in the qualia structure must be fully interpreted in the resulting syntactic structure (cf. Chomsky (1981)):

(38) QUALIA SATURATION: A qualia structure is saturated only if all arguments in the qualia roles are *covered*.

(39) COVERING:
An argument x is covered only if:
(i) x is linked to a position in s-structure; or
(ii) x is logically dependent on a covered argument y; or
(iii) x is existentially closed by virtue of its type.

Pustejovsky (1995:103)

We assume that the condition stated in (39)iii refers to incorporated arguments, i.e., shadow type arguments, that by virtue of their type (parameters incorporated in the lexical item, different from proper arguments, that can only be syntactically expressed by subtyping or specification processes) are not necessarily realized in s-structure. Logically dependent arguments (in (39)ii) are, for instance, arguments that are recovered from the qualia structure of a covered argument, linked to a position in s-structure, as the case of the default argument of the verb *build* illustrates (see (28) above). However, and to account for the fact that the second argument in the semantic predicate stated in the Agentive role of the verb *build* (Agentive = build_act(e_1, x:human, z: material)) is not projected into an object position, as predicted in (35), Pustejovsky (1995:104) postulates that "such a default argument can be viewed as a Skolem function of the argument it is dependent on, namely *f(y)*. Thus, given the calculus of relations in the qualia and the templates associated with them, the FORMAL argument in this case ends up bound to the object position in syntax [Q_A: R(e_1, x, f(y))$\rightarrow$ x: SUBJ, y: OBJ]".

This resolution can lead to the assumption that default arguments are syntactically realized in oblique position, mediated by a function indicator item, such as prepositions for instance. However, this assumption lacks empirical motivation. If, on the one hand, default arguments are usually realized as prepositional phrases, there is no solid reason, on the other, to assume that this is always the case. Some verbs select true arguments that are realized by prepositional phrases instead of projecting an object position (for example, *dream* ***of***, *gostar* ***de*** (≅like **of**), etc.), whereas some arguments occurring in object position may not correspond to true arguments of the verb (for instance, *run* ***the marathon***, *sleep* ***the whole night***). That is, not always selection properties and subcategorization properties are uniformly mapped. As it was already mentioned in section 4.1.1.1 above, argument structures can contemplate subcategorization constraints through the recursive use of available structures in the lexicon, as exemplified in (14), allowing to account for the idiosyncratic subcategorization properties of lexical items and enabling, at the same time, the correct projection of the semantic predicates stated in the qualia structure of verbs and nouns.

The second issue related to qualia saturation concerns the information stated in the lexical entries of nominal items. Qualia saturation, as formulated above, cannot be extended to the qualia structure of nouns, since the arguments in nominal qualia structures are not necessarily projected into syntactic form, with the exception of event denoting nouns. In fact, it is only when the lexical item denotes an event, thus having an event structure associated to it, that qualia saturation applies, even if not exactly in the terms formulated above. In fact, and as stated by Pustejovsky (1995:191), "only arguments associated with the head event are obligatorily expressed at surface structure". To provide a universal mapping rule for all event denoting POS, we propose the statement of this restriction within the qualia saturation principle resulting in the event structure saturation principle, which can be formulated as follows:

(40) a. The values of the event structure must be *saturated* at syntax.

b. EVENT STRUCTURE SATURATION: An event structure is saturated only if all the arguments of the semantic predicate the head event, expressed in the qualia roles, are *covered*.

(41) COVERING:

An argument x is covered only if:

(i) x is linked to a position in s-structure; or

(ii) x is logically dependent on a covered argument y; or

(iii) x is existentially closed by virtue of its type.

This way it is possible to account for argument mapping of event denoting nouns, such as *burning*, in (42), adjectives (see (43)), and verbs (like *sink* (see (44)) with the same rule.

(42) burning

$$\begin{bmatrix} \text{ARGSTR} = \begin{bmatrix} \text{P-ARG}_1 = \text{e}_1 \\ \text{ARG}_1 = \text{x:physical object} \end{bmatrix} \\ \\ \text{EVENTSTR} = \begin{bmatrix} \text{E}_1 = \text{e}_1\text{: process} \\ \text{HEAD} = \text{e}_1 \end{bmatrix} \\ \text{QUALIA} = [\text{FORMAL} = \text{burn\_act}(\text{e}_1, \text{x})] \end{bmatrix}$$

(43) capable

$$\begin{bmatrix} \text{ARGSTR} = \begin{bmatrix} \text{P-ARG}_1 = \text{e}_1 \\ \text{ARG}_1 = [\text{x : individual}] \\ \\ \text{ARG}_2 = \begin{bmatrix} of \\ \text{ARGSTR} = [\text{ARG}_1 = \text{y : event}] \\ \ldots \end{bmatrix} \end{bmatrix} \\ \text{EVENTSTR} = \begin{bmatrix} \text{E}_1 = \text{e}_1\text{: state} \\ \text{HEAD} = \text{e}_1 \end{bmatrix} \\ \\ \text{QUALIA} = [\text{FORMAL} = \text{capable}(\text{e}_1, \text{x}, \text{y})] \end{bmatrix}$$

(44) sink

$$\begin{bmatrix} \text{ARGSTR} = \begin{bmatrix} \text{P-ARG}_1 = \text{e}_1 \\ \text{ARG}_1 = \text{x : entity} \\ \text{ARG}_2 = \text{y : phys\_object} \end{bmatrix} \\ \\ \text{EVENTSTR} = \begin{bmatrix} \text{E}_1 = \text{e}_1\text{: process} \\ \text{E}_2 = \text{e}_2\text{: state} \\ \text{RESTR} = <\alpha \\ \text{HEAD} = \end{bmatrix} \\ \\ \text{QUALIA} = \begin{bmatrix} \text{FORMAL} = \neg\text{under\_water}(\text{e}_2, \text{y}), \text{under\_water}(\text{e}_2, \text{y}) \\ \text{AGENTIVE} = \text{sink\_act}(\text{e}_1, \text{x}, \text{y}) \end{bmatrix} \\ \end{bmatrix}$$

Event structure saturation assures the correct mapping from the semantic predicates expressed in the qualia structure, given the subevents considered in the event structure as well as the prominent event projected into the s-structure, while the argument structure determines the number, semantic type, but also the syntactic expression of the arguments of a given lexical item, regardless of its POS. In the cases of atomic events (i.e., when the event structure is composed of only one event, as in (42) and (43)), the head event corresponds to the only event in the event structure. In the case of transition denoting verbs, as in (44), the non specification of the HEAD feature accounts for the different syntactic projections dependent on the head event instantiated in context.

Together, the three levels of representation – argument structure (stating the number, semantic type and subcategorization properties of the arguments of lexical items), event structure (determining the type, number, relation and head subevents of the event denoted by lexical items), and qualia structure (expressing the semantic predicates that establish the relations that characterize the meaning of lexical items) – allow the description of the semantic content of lexical items, also providing syntactic mapping information.

Note that there are some unaddressed mapping issues, namely in what concerns the syntactic realizations of arguments in alternative constructions. For instance, in non-causative constructions, the second argument in the argument structure is syntactically realized in subject position, whereas the first argument, when realized, is expressed by a prepositional phrase. In other words, the verb argument structure in inchoative constructions consists of a true argument and of a default argument, whereas in causative constructions it consists of two true arguments. This implies that, at the lexical level, the syntactic mapping information necessarily concerns the basic position of arguments, not accounting for all the possible syntactic positions.

4.1.1.4 Lexical inheritance structure

The lexical inheritance structure is the fourth level of representation considered in the GL model. Within this model, in order to fully describe the semantics of a lexical item it is necessary to link it to a semantic type integrated in a type lattice, allowing the inheritance of features from its mother node(s). Lexical inheritance structure is, thus, the level of representation that allows the access to the semantic types, organized in an orthogonal type lattice.

To regulate multiple inheritance, that is, the mechanism that allows the relative prominence of different aspects of a given object (see (45)), Pustejovsky (1995) proposes that lexical inheritance is accomplished through qualia vectors: the nodes in the type lattice are linked according to qualia values (in (46)). In this way it is possible to assure that only the correct inferences are generated, accounting for different behaviors of sister nodes such as *dictionary* and *novel*.

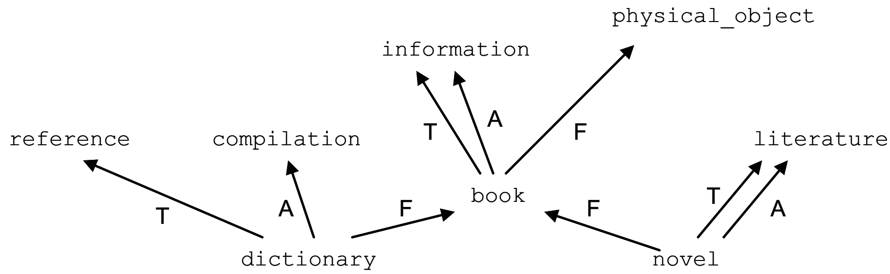
(45) a. #John read the dictionary.
b. John read the novel.
c. John consulted the dictionary.
d. #John consulted the novel.

(46) a. *book* is_Formal `physical_object`
book is_Telic `information`
book is_Agentive `information`

b. *dictionary* is_Formal `book`
dictionary is_Telic `reference`
dictionary is_Agentive `compilation`

c. *novel* is_Formal `book`
novel is_Telic `literature`
novel is_Agentive `literature`

(47) Type lattice for *book*, *dictionary* and *novel*



Adapted from Pustejovsky(1995: 144-145)

The association of a given lexical item to its semantic type is represented by the direct association of the argument to its semantic type in argument structure, example: ARGSTR = [ARG_1 = x: `book`] (in our proposal, this would be formalized as the value of proper arguments, ARGSTR = [P-ARG_1 = x: `book`]), or by the Formal role value in the qualia structure, example: Q = [FORMAL = book(x). Moreover, the semantic predicates in the qualia structure represent the constraints on the type of *book* considered. For instance, Q = [FORMAL = book(x); AG = compile(e_1, y, x); TELIC = consult(e_2, y, x)], represents the noun *dictionary*.

In our perspective, the formalization of the information on this level of representation raises at least two issues. On the one hand, contrary to the values set for most noun qualia roles, the semantic predicates in the qualia structures of event denoting lexical items do not mirror the qualia vectors regulating the inherited information from mother nodes within the type lattice. On the other, the relation between the semantic object denoted by a given lexical item and its semantic type, established at the Formal role, is not parallel to the relations established through semantic predicates in the other qualia roles. Specifically, and considering the noun *dictionary* again, the relation established through the value of the Telic role (consult(e_2, y, x)) only indirectly refers to the semantic type `reference`, and through mechanisms not described by Pustejovsky (1995).

(48) novel

$$\begin{bmatrix} \text{ARGSTR} = \begin{bmatrix} \text{P-ARG}_1 = \text{x}:\text{book} \end{bmatrix} \\ \\ \text{QUALIA} = \begin{bmatrix} \text{FORMAL} = \text{book(x)} \\ \text{TELIC} = \text{consult(e}_1\text{, y:human, x)} \\ \text{AGENTIVE} = \text{compile(e}_2\text{, w:lexicographer, x)} \end{bmatrix} \\ \\ \end{bmatrix}$$

For these reasons, and given our goal to incorporate GL structures in WordNet.PT, we propose an approach making use of the conceptual relations already established within the wordnet model. This way, lexical inheritance information is stated through the subtyping relation 'x: y → x is hyponym of y' at the argument structure level, as a value of proper-arguments.

(49) dictionary

$$\begin{bmatrix} \text{ARGSTR} = \begin{bmatrix} \text{P-ARG}_1 = \text{x}: \begin{bmatrix} \text{book} \\ \text{ARGSTR} = \begin{bmatrix} \text{P-ARG}_1 = \text{x}:\text{artifact} \\ \text{P-ARG}_2 = \text{y}:\text{information} \end{bmatrix} \\ \\ \text{QUALIA} = \begin{bmatrix} \text{FORMAL} = \boxed{1}\text{contains(e}_1\text{, x, y)} \\ \text{CONST} = \boxed{2}\text{has\_part(x} \bullet \text{y,} \begin{bmatrix} \text{page} \\ \ldots \end{bmatrix}\text{),} \\ \qquad \text{has\_part(x} \bullet \text{y,} \begin{bmatrix} \text{cover} \\ \ldots \end{bmatrix}\text{)} \\ \text{TELIC} = \text{read(e}_2\text{,} \begin{bmatrix} \text{human} \\ \ldots \end{bmatrix}\text{, x} \bullet \text{y)} \\ \text{AGENTIVE} = \text{write(e}_3\text{,} \begin{bmatrix} \text{human} \\ \ldots \end{bmatrix}\text{, x} \bullet \text{y)} \end{bmatrix} \end{bmatrix} \end{bmatrix} \\ \\ \text{QUALIA} = \begin{bmatrix} \text{FORMAL} = \boxed{1} \\ \text{CONST} = \boxed{2} \\ \text{TELIC} = \text{consult(e}_1\text{,} \begin{bmatrix} \text{human} \\ \ldots \end{bmatrix}\text{, x)} \\ \text{AGENTIVE} = \text{compile(e}_2\text{,} \begin{bmatrix} \text{lexicographer} \\ \ldots \end{bmatrix}\text{, x)} \end{bmatrix} \end{bmatrix}$$

Proper-arguments are, thus, linked to their superordinate in the hierarchy, having access to its informational structure, as depicted in (49) above.

We argued that this proposal has several advantages. First, it accurately distinguishes the subtyping relation established between the proper argument and its mother node from the characterization of the variables used in the semantic predicates of qualia roles. In fact, what is intended to be represented is that a *novel* is a subtype of *book*, but not that the entities that are part of *book* are subtypes of *page* or *cover*, for instance. A *book* can have a *paperback*, which is a subtype of *cover*, but the semantic predicate in its Constitutive role does not stand for "a book has as part any subtype of cover", rather it stands for "a book has as part a cover or any of its subtypes". Second, it allows to straightforwardly establishing what is inherited from the mother node and what constitues the specific contribution of a lexical item through the direct access to the content of the hyperonym node and the use of available informational structures in the lexicon, preserving, nonetheless the qualia vectors that regulate multiple inheritance. This strategy also contributes to the coherence of the model in what concerns the Formal role values established for nouns and verbs. The definition of lexical-conceptual subtypes being established at the argument structure, the Formal role remains available to establish stative properties associated to a given object: state of affairs for event denoting items; the relation between the proper-arguments of regular polysemous items, as well as properties that distinguish a given object within its semantic domain (in this case, that distinguish it from its hyperonym), such as FORMAL = round(e_1,x) for the noun *ball*, for instance.

This proposal follows directly from the Lexical Inheritance Structure proposed by Pustejovsky (1995) in what concerns the characterization of lexical items through the use of conceptual types, but profits from the WordNet model organization which already establishes lexical conceptual relations. This way, it is not necessary to build a semantic type hierarchy to describe the semantics of lexical items, since it is possible to use the lexical structures available in the lexical-conceptual hierarchy.

Considering the nodes in the lexicon as the semantic types proceeds from the fact that, as a model of the mental lexicon, WN can be seen as a reflex of the relation between lexical knowledge and our ability to organize and classify the concepts in the world (see Pustejovsky (2001)). As a consequence we would have that, when referring to an argument of type `human`, we would access not the type and the relations established between this type and the other types in the hierarchy of types, but the synset $\{human\}_N$ and the relations established between this node and all the other nodes in the lexicon. Phenomena such as regular polysemy for instance, covered by complex types such as `information•physical object`[4] (as in (48) and (49) above), are already accounted for in WN by considering that $\{book\}_N$ has two hyperonyms, $\{information\}_N$ and $\{artifact\}_N$.

Also, the hyperonym relation enables lexical inheritance through the use of available lexical structures. All that is required is that the semantic specification of the arguments and the values of the Qualia roles in a lexical entry do not refer to types pertaining to an extra-lexicon hierarchy, but to other lexical items within the lexicon. Note that lexical gaps (concepts that are

[4] The semantic operator '•' is used to represented complex type objects, that is, lexical expressions that exhibit regular polysemy, such as *book*. The operators used in GL to model these expressions are '•' – when both types of the complex object can be simultaneously interpreted, `information•physical object`, `opening•physical object`, for instance –, and '∘' (Buitelaar 1998) when the types that compose the complex object cannot be simultaneously interpreted, like `animal∘food`.

not lexicalized in a given language) might raise important issues to this proposal. However, and besides seeming reasonable to consider that the ontological and hierarchical nature of a hand-build relational lexicon should be enough to ensure the conceptual categorization needed to describe the meaning of lexical items, the WN model already considers lexical gaps, in what concerns equivalence relations between languages for instance.

Given these considerations, for the purposes of the present work, a lexical item, α, is represented by information in three distinct levels – argument structure (A), event structure (E) and qualia structure (Q) – and integrated in a lexical inheritance structure (I), as reformulated here in (50).

(50) $\alpha = \langle A, E, Q \rangle \in I$

4.1.2 Generative mechanisms

The GL model, besides assuming a unification device that allows for information sharing between structures, represented by co-indexed values in the AVM structures, defines three major generative mechanisms – Co-composition, Selective Binding and Type Coercion – that capture "the means by which words can assume a potentially infinite number of senses in context, while limiting the number of senses actually stored in the lexicon." (Pustejovsky 1995: 105).

The action of these mechanisms is directly conditioned by lexical government and binding, in the sense that a syntactic structure cannot be interpreted outside a semantic and syntactic context. Thus, through the connection between the several levels of semantic information in lexical entries, the GL model provides a simple and accurate way to account for meaning in context and for the polymorphic behavior of lexical items.

In this subsection we will briefly present these generative mechanisms, illustrating how they interact with the representation levels considered.

4.1.2.1 Co-composition

Co-composition, as the name indicates, comprises the semantic operations that allow for the completion of semantically underspecified forms with information present in the semantic content of its arguments. Included in this set are manner co-composition, feature transcription or light verb specification, for instance. Based on Keenan & Faltz (1985), co-composition is defined as follows:

(51) "For two expressions α, of type <a,b>, and β, of type a, with qualia structures QS_{α} and QS_{β}, respectively, then, if there is a quale value shared by α and β, $[QS_{\alpha} \ldots [Q_i = \gamma]]$ and $[QS_{\beta} \ldots [Q_i = \gamma]]$, then we can define the qualia unification of QS_{α} and QS_{β}, $QS_{\alpha} \cdot\!\cap\, QS_{\beta}$, as the unique greatest lower bound of these two qualia structures. Further, $\alpha(\beta)$ is of type b with $QS_{\alpha(\beta)} = QS_{\alpha} \cap QS_{\beta}$"

Pustejovsky (1995: 124)

As expressed by this formulation, co-composition results in the qualia unification of both expressions. To illustrate this mechanism, Pustejovsky uses the verbs *bake* and *float* (Pustejovsky 1995: 122, 125-126), which are differently interpreted depending on the arguments selected, in (52)a and b, or on the construction in which the verb occurs, in (53)a and b.

(52) a. John baked the potatoes. – *change of state*

b. John baked the cake. – *creation*[5]

(53) a. The bottle is floating in the river. – *process*

b. The bottle floated into the cave. – *transition*

Assuming that there is a single lexical entry for the verb *float,* for instance, and considering the representation for the PP *into the cave* (adapted from Pustejovsky 1995: 126), in (54) and (55), respectively, it is possible to generate both interpretations through co-composition (see (56)).

[5] The resulting AVM for the VP *bake a cake* proposed by Pustejovsky (1995:125), presented below, is not considered here since it involves several issues not address by the author. For instance, the definition of the head event, whose semantic predicate in the qualia structure defined does not correspond to the projected syntactic form: *bake a cake* and not *bake a mass*.

$$
\begin{bmatrix}
\text{bake a cake} & \\
\text{EVENTSTR} = & \begin{bmatrix} E_1 = e_1 : \text{process} \\ E_2 = e_2 : \text{state} \\ \text{RESTR} = <\alpha \\ \text{HEAD} = e_1 \end{bmatrix} \\
\text{ARGSTR} = & \begin{bmatrix}
\text{ARG}_1 = \boxed{1} \begin{bmatrix} \text{animate\_ind} \\ \text{FORMAL} = \text{phyobj} \end{bmatrix} \\
\text{ARG}_2 = \boxed{2} \begin{bmatrix} \text{artifact} \\ \text{CONST} = \boxed{3} \\ \text{FORMAL} = \text{phyobj} \end{bmatrix} \\
\text{D-ARG}_1 = \boxed{3} \begin{bmatrix} \text{material} \\ \text{FORMAL} = \text{mass} \end{bmatrix}
\end{bmatrix} \\
\text{QUALIA} = & \begin{bmatrix} \text{create-lcp} \\ \text{FORMAL} = \text{exist}(e_2, \boxed{2}) \\ \text{AGENTIVE} = \text{bake\_act}(e_1, \boxed{1}, \boxed{3}) \end{bmatrix}
\end{bmatrix}
$$

(54) float

$$\begin{bmatrix} \text{ARGSTR} = \begin{bmatrix} \text{P - ARG1} = \text{e1} \\ \text{ARG1} = \text{x} = \begin{bmatrix} \text{physical object} \\ \ldots \end{bmatrix} \\ \text{D} - \text{ARG1} = \text{y} = \begin{bmatrix} \text{in} \\ \text{ARG1} = \text{x} = \begin{bmatrix} \text{water} \\ \ldots \end{bmatrix} \\ \ldots \end{bmatrix} \end{bmatrix} \\ \text{EVENTSTR} = \begin{bmatrix} \text{E1} = \text{e2} : \text{process} \\ \text{HEAD} = \text{e2} \end{bmatrix} \\ \text{QUALIA} = [\text{FORMAL} = \text{float(e3, x, y)}] \end{bmatrix}$$

(55) into the cave

$$\begin{bmatrix} \text{ARGSTR} = \begin{bmatrix} \text{P - ARG1} = \text{x} : \begin{bmatrix} \text{to} \\ \ldots . \end{bmatrix} \\ \text{ARG1} = \text{y} = \begin{bmatrix} \text{physical object} \\ \ldots \end{bmatrix} \\ \text{ARG2} = \text{z} = \begin{bmatrix} \text{cave} \\ \ldots \end{bmatrix} \end{bmatrix} \\ \text{EVENTSTR} = \begin{bmatrix} \text{E1} = \text{e2} : \text{state} \\ \text{RESTR} = < \alpha \\ \text{HEAD} = \text{e2} \end{bmatrix} \\ \text{QUALIA} = [\text{FORMAL} = \neg\text{at(e2, y, z), at(e2, y, z)}] \end{bmatrix}$$

As it is possible to see in (56), and as formulated in (51) above, co-composition generally corresponds to qualia unification. In this case, the Formal value of the PP corresponds to the Formal value of the VP, represented here by the numerical index 3, the argument 1 of both predicates being also unified, represented here by the numerical index 1. The resulting event structure of the VP is constituted by the events denoted both by the verb *float* and by the PP *into the cave* (numerical index 2).

(56) floated into the cave

$$\begin{bmatrix} \text{ARGSTR} = \begin{bmatrix} \text{P-ARG1} = \text{e1} \\ \text{ARG1} = \text{x} = \boxed{1}\begin{bmatrix}\text{physical entity} \\ \ldots\end{bmatrix} \\ \text{ARG2} = \text{y} = \begin{bmatrix} \text{into} \\ \text{ARGSTR} = \begin{bmatrix} \text{P-ARG1} = \text{e}: \begin{bmatrix}\text{to} \\ \ldots.\end{bmatrix} \\ \text{ARG1} = \text{y} = \boxed{1} \\ \text{ARG2} = \text{z} = \begin{bmatrix}\text{cave} \\ \ldots\end{bmatrix} \end{bmatrix} \\ \text{EVENTSTR} = \begin{bmatrix} \text{E1} = \boxed{2}\,\text{e2}: \text{state} \\ \text{RESTR} = <\alpha \\ \text{HEAD} = \text{e2} \end{bmatrix} \\ \text{QUALIA} = \begin{bmatrix}\text{FORMAL} = \boxed{3}\,\neg\text{at}(\text{e1}, \text{x}, \text{y}), \text{at}(\text{e2}, \text{x}, \text{y})\end{bmatrix} \end{bmatrix} \\ \text{D}-\text{ARG1} = \text{w} = \begin{bmatrix} \text{in} \\ \text{ARG1} = \text{x} = \begin{bmatrix}\text{water} \\ \ldots\end{bmatrix} \\ \ldots \end{bmatrix} \end{bmatrix} \\ \text{EVENTSTR} = \begin{bmatrix} \text{E1} = \text{e1}: \text{process} \\ \text{E2} = \boxed{2} \\ \text{RESTR} = <\alpha \\ \text{HEAD} = \boxed{2} \end{bmatrix} \\ \text{QUALIA} = \begin{bmatrix}\text{FORMAL} = \text{float}(\text{e1}, \boxed{1}, \text{w}), \boxed{3}\end{bmatrix} \end{bmatrix}$$

Thus, co-composition allows the determination of meaning in context without resorting to multiple entries in the lexicon.

4.1.2.2 Selective binding

Selective binding is the generative mechanism that concerns the description of the relation between a modifier and the modified item. As indicated by its designation, selective binding allows the modifier to select its arguments from the set of objects in the semantic content of the modified item. Selective binding is, thus, formalized as follows:

(57) "If α is of type <a,a>, β is of type b, and the qualia structure of β, QS_β, has a quale q of type a, then $\alpha\beta$ is of type b where $\|\alpha\beta\| = \beta \cap \alpha(q_\beta)$"

(Pustejovsky 1995: 129)

This generative mechanism allows for accounting for context sensibility of adjectives such as *fast*, depending on the modified noun, given that the resulting interpretations emerge from the modification of the selected qualia role value present in the semantic content of the noun:

(58) a. a fast car ≅ a car that rides fast

b. a fast typist ≅ a typist that types fast

$$\begin{bmatrix} \text{a fast car/typist} & \\ \text{ARGSTR} = & \begin{bmatrix} \text{P-ARG}_1 = \text{e}_1 \\ \text{ARG}_1 = \text{x} = \begin{bmatrix} \text{car/typist} \\ \ldots \\ \text{QUALIA} = \left[\textbf{TELIC} = \boxed{1}\,\textbf{rides(e}_1\textbf{, x)/ types(e}_1\textbf{, x)}\right] \end{bmatrix} \end{bmatrix} \\ \text{EVENTSTR} = & \begin{bmatrix} \text{E}_1 = \text{e}_2 : \text{state} \\ \text{HEAD} = \text{e}_2 \end{bmatrix} \\ \text{QUALIA} = & \left[\text{FORMAL} = \textbf{fast(e}_2\textbf{,}\,\boxed{1}\textbf{)}\right] \end{bmatrix}$$

The selective binding mechanism accounts for the polysemic behavior of adjectives such as *good* and *fast*, by establishing that these adjectives select for the Telic role of the modified noun. Many adjectives select a given qualia role value of the noun they modify (for instance, adjectives such as *big* or *round* should select for Formal values), selective binding being in fact a decisive mechanism to represent and predict the behavior of modifiers, also without resorting to multiple entries that contemplate all the meaning variations possible.

4.1.2.3 Type coercion

The third generative mechanism considered in GL is type coercion. This mechanism consists in a semantic operation that converts an argument of a given type into the type expected by the predicate. Although type coercion may reflect on the syntactic expression of the predicate, this semantic operation does not involve changes to the syntactic type of the item, but rather to its semantic type, which may be associated to different canonical syntactic forms. Type coercion is formulated by Pustejovsky (1995:111) as follows. Being Σ_α a set of shifting operators which can operate over a given expression α:

(59) If α is of type c, and β is of type <a,b>, then

(i) if type c = a, then $\beta(\alpha)$ is of type b.

(ii) If there is a $\sigma \in \Sigma_\alpha$ such that $\sigma(\alpha)$ results in an expression of type a, then $\beta(\sigma(\alpha))$ is of type b.

(iii) Otherwise a type error is produced.

Let us consider, again, the verb *enjoy*, whose argument structure is presented in (60).

(60) a. enjoy

$$\begin{bmatrix} \text{ARGSTR} = \begin{bmatrix} \text{P-ARG}_1 = e_1 \\ \text{ARG}_1 = x = [\text{individual}] \\ \text{ARG}_2 = y = [\text{event}] \end{bmatrix} \\ \ldots \end{bmatrix}$$

b. She enjoys [movies]$_{\text{individual}}$.

Although the verb *enjoy* selects an event type argument (ARG_2 = y = [event]), it can, nonetheless, occur with an individual type argument, as illustrated by (60)b. This is possible through type coercion, since it is possible to "extract" an expression corresponding to the type expected by the predicate from the semantic content of the noun *movie*, namely the value of its telic role:

(61) movie

$$\begin{bmatrix} \text{ARGSTR} = [\text{P-ARG}_1 = x] \\ \text{QUALIA} = \begin{bmatrix} \ldots \\ \text{TELIC} = \text{watch}(e_1, [\text{human}], x) \end{bmatrix} \end{bmatrix}$$

The simplest shifting operator included in the set of operators considered for the application of type coercion is the subtyping operator. Subtyping allows the prediction of well-formed expressions when a subtype of the type expected by the predicate occurs:

(62) John drives a vehicle/car/pickup/Ford to work.

The general case described in the semantic content of the verb *drive* is that this predicate selects for a vehicle type argument. However, and given the percolation of information within a hierarchy of type – a subtype being all its supertype is plus more – this argument can be realized by the type `vehicle` or by any of its subtypes.

Note that the replacement of the semantic type lattice used in GL for the available entries in a wordnet does not affect the application of type coercion in any way, since the position in the net corresponds to a lexical-conceptual node similar to a semantic type. Also, type coercion, as all the other generative mechanisms in GL, is necessarily conditioned by the information stated in the qualia structure of the lexical items, a level of representation entirely preserved within our approach.

As a final remark regarding the generative mechanisms considered in the GL model, we want to stress that, allowing for underspecified lexical representations, these semantic mechanisms operate at phrase level and can be considered to explain the polymorphic behavior of the verbs in study. Nonetheless, at lexical entry level, only subtyping is overtly considered in stating the arguments selected by Portuguese movement verbs.

4.2 In summary

After this presentation and discussion of the GL model, and given the reformulations proposed in this chapter, in this section we summarize the structures and definitions assumed in the remainder of the work depicted in this dissertation for the representation of Portuguese verbs of movement, integrated in WN.PT.

A lexical item, α, is represented by the information in argument structure (A), event structure (E) and qualia structure (Q) levels, and integrated in a lexical-conceptual relational lexicon (I):

(63) $\alpha = \langle A, E, Q \rangle \in I$

Argument structure (ARGSTR): level of representation in which the number and type of arguments of a lexical item is stated, contemplating the definition of the semantic properties of the logical arguments of a given lexical item, as well as syntactic mapping information. The order by which arguments are stated within the argument structure expresses constraints on the syntactic mapping: arguments are listed from the less oblique position to the more oblique one. Lexical items that introduce oblique arguments are overtly stated through the use of the available structures in the lexicon. There are four types of arguments considered in the argument structure of lexical items:

i) Proper arguments (P-ARG_n): parameters of the lexical item semantic content that correspond to the semantic object denoted by the lexical item.

ii) True arguments (ARG_n): parameters of the lexical item semantic content whose syntactic omission can only be licensed when recoverable from context.

iii) Default arguments (D-ARG_n): parameters required by the logical expressions in the qualia structure of the lexical item that can be syntactically expressed by subtyping or specification processes.

iv) Shadow arguments (S-ARG_n): parameters incorporated in the lexical item that can only be syntactically expressed by subtyping or specification processes.

True arguments, default arguments and shadow arguments variables are associated to the node in the hierarchical lexicon that defines their syntactic expression: any argument can be realized by the lexical expression to which it is associated, or by any of its hyponyms.

Event structure (EVENTSTR): level of representation which refers to the properties of an event associated to a lexical item. It is constituted by the event arguments (E_1, …, E_n) corresponding to the subevents of a given event (corresponding to the value of the proper argument) and their respective Aktionsart type, by the attribute Restrictions (RESTR), whose values represent the temporal order constraints holding between the subevent arguments, and by the attribute Head (HEAD) which allows the representation of the head subevent. There are three types of event arguments:

i) States: atomic events, non evaluated regarding any other, i.e., given a time interval I, a state is an event that is verified in all I: $[\text{EVENTSTR} = [\text{E}1 = \text{e}1 : \text{state}]]$

ii) Processes: sequences of identical events, i.e., given a time interval I, a process is a sequence of events that is verified in all the intervals of I: $[\text{EVENTSTR} = [\text{E}1 = \text{e}1 : \text{process}]]$

iii) Transitions: events evaluated regarding another event, i.e., in a time interval I, a transition begins with a process that leads to a final state, different from the initial one:

$$\begin{bmatrix} \text{EVENTSTR} = \begin{bmatrix} \text{E}1 = \text{e}1 : \text{process} \\ \text{E}2 = \text{e}2 : \text{state} \\ \text{RESTR} = <\alpha \end{bmatrix} \end{bmatrix}$$

The values of the Restrictions attribute refer to three possible temporal ordering relations:

i) Exhaustive ordered part of ($<_\alpha$): e_3 is a complex event constituted by e_1 and e_2 subevents, logical parts of e_3, e_1 precedes e_2, and e_1 and e_2 are the only subevents of e_3:

a. $[_{e3}\ e_1 <_\alpha e_2] =_{def} <_\alpha (\{e_1, e_2\}, e_3)$

b. $\forall e_1, e_2, e_3\ [<_\alpha (\{e_1, e_2\}, e_3\}) \leftrightarrow e_1 \preccurlyeq e_3 \wedge e_2 \preccurlyeq e_3 \wedge e_1 < e_2 \wedge\ \forall e[e \preccurlyeq e_3 \rightarrow e = e_1 \vee e_2]]$

ii) Exhaustive overlap part of ($\circ_\alpha$,): e_3 is a complex event constituted by e_1 and e_2 subevents, logical parts of e_3, e_1 temporally includes e_2, e_2 temporally includes e_1, both e_1 and e_2 being temporally included in e_3, and e_1 and e_2 being the only subevents of e_3:

a. $[_{e3}\ e_1 \circ_\alpha e_2] =_{def} \circ_\alpha (\{e_1, e_2\}, e_3)$

b. $\forall e_1, e_2, e_3\ [\circ_\alpha (\{e_1, e_2\}, e_3\}) \leftrightarrow e_1 \preccurlyeq e_3 \wedge e_2 \preccurlyeq e_3 \wedge e_1 \sqsubseteq e_2 \wedge e_2 \sqsubseteq e_1 \wedge \exists e[e \sqsubseteq e_1 \wedge e \sqsubseteq e_2 \wedge e = e_3] \wedge \forall e[e \preccurlyeq e_3 \rightarrow e = e_1 \vee e_2]]$

iii) Exhaustive ordered overlap ($<\circ_\alpha$,): e_3 is a complex event constituted by e_1 and e_2 subevents, logical parts of e_3, e_1 temporally includes e_2, the initial part of e_1 precedes the initial part of e_2, the ending parts of e_1 and e_2 being simultaneous, and e_1 and e_2 being the only subevents of e_3:

a. $[_{e3}\ e_1 <\circ_\alpha e_2] =_{def} <\circ_\alpha (\{e_1, e_2\}, e_3)$

b. $\forall e_1, e_2, e_3\ [<\circ_\alpha (\{e_1, e_2\}, e_3\}) \leftrightarrow e_1 \preccurlyeq e_3 \wedge e_2 \preccurlyeq e_3 \wedge e_1 \circ e_2 \wedge \text{init}(e_1) < \text{init}(e_2) \wedge \text{end}(e_1) = \text{end}(e_2) \wedge \forall e[e \preccurlyeq e_3 \rightarrow e = e_1 \vee e_2]]$

The attribute Head enables the statement of the prominent subevent in an event structure, which is projected into syntax, the values of the event structure being necessarily *saturated* by syntax.

(64) EVENT STRUCTURE SATURATION: An event structure is saturated only if all the arguments of the head event semantic predicate, expressed in the qualia roles, are *covered*.

(65) COVERING:
An argument x is covered only if:
(i) x is linked to a position in s-structure; or
(ii) x is logically dependent on a covered argument y; or
(iii) x is existentially closed by virtue of its type.

Qualia structure (QUALIA): level of representation that provides the semantic objects that define the meaning of a lexical item. It is composed of four attributes or qualia roles:

i) Constitutive role (CONST): attribute whose values express the relation between a given object and its constituents or parts, such as material, weight or the semantic predicates that form the sequence of identical events that constitutes a process.

ii) Formal role (FORMAL): attribute whose values establish the stative properties that distinguish a given object within its semantic domain, such as the sortal typing of the proper-argument, the relation between proper-arguments of different types, or semantic predicates corresponding to actual state of affairs, typically states.

iii) Telic role (TELIC): attribute whose values concern the information about the function or purpose of the object or of the event.

iv) Agentive role (AGENTIVE): attribute whose values determine the origin or the causal chain involved in the bringing about of the object or of the event.

The values of qualia roles are stated through semantic predicates, with well-defined types and relational structures, indicating the proper binding of the predicating term. In what concerns predicates, qualia role values are mapped to syntax according to the argument types defined and through the following mapping rules:

(66) a. Q_i: R(e_1, x, y) $\rightarrow$ x:SUBJ, y:OBJ

b. Q_j: P(e_2, y) $\rightarrow$ y:SUBJ

Lexical-conceptual relational lexicon (I): representation level that assures the hierarchical organization of lexical items and provides the relations between a given lexical structure and other lexical structures in the lexicon. Lexical inheritance information is stated through the subtyping relation '*x: y → x is hyponym of y*' at argument structure level. Proper-arguments are, thus, linked to their superordinates in the hierarchy, having access to their informational structure. Also, other relations are regulated through the relations established at qualia structure level.

Considering that a semantic type corresponds to a position in the lexical-conceptual relational lexicon, the generative mechanisms that allow the interaction between lexical structures can be defined as follows:

i) Co-composition:

(67) "For two expressions α, of type <a,b>, and β, of type a, with qualia structures QS_α and QS_β, respectively, then, if there is a quale value shared by α and β, $[QS_\alpha \ldots [Q_i = \gamma]]$ and $[QS_\beta \ldots [Q_i = \gamma]]$, then we can define the qualia unification of QS_α and QS_β, $QS_\alpha \cdot\!\cap QS_\beta$, as the unique greatest lower bound of these two qualia structures. Further, $\alpha(\beta)$ is of type b with $QS_{\alpha(\beta)} = QS_\alpha \cap QS_\beta$"

(Pustejovsky 1995: 121)

ii) Selective binding:

(68) "If α is of type <a,a>, β is of type b, and the qualia structure of β, QS_β, has a quale q of type a, then $\alpha\beta$ is of type b where $//\alpha\beta// = \beta \cap \alpha(q_\beta)$"

(Pustejovsky 1995: 129)

iii) Type coercion: being Σ_α a set of shifting operators wich can operate over a given expression α

(69) "If α is of type c, and β is of type <a,b>, then

(i) if type c = a, then $\beta(\alpha)$ is of type b.

(ii) If there is a $\sigma \in \Sigma_\alpha$ such that $\sigma(\alpha)$ results in an expression of type a, then $\beta(\sigma(\alpha))$ is of type b.

(iii) Otherwise a type error is produced."

Pustejovsky (1995:111)

4.3 Conclusions

In this chapter we focused on the GL model, with special attention to the modeling of verbs. Given the goals in pursue, namely the integration of GL lexical structures in wordnets, we discussed several aspects related to the integration of its levels of representation on a lexical-conceptual relational lexicon – considering the information already established in these lexica –, but also in what concerned the systematicity and consistency of the model for all POS. This way, the redefinitions proposed serve two purposes: the enhancement of the accuracy and coherence of the GL model and the combination of the GL and WordNet models.

This discussion allowed us to observe the compatibility of the two models, given the similarities of the information stated at lexical level (part of relations, function relations, and so on), on the one hand, and the almost perfect juxtaposition of the two models (wordnet can provide a lattice in which GL lexical inheritance structure is based), on the other. Also, and given that GL is, in fact, a qualia based lexical model, we can benefit as well from current research on the integration of qualia information in the WordNet model, for co-hyponym differentiation purposes, for instance, by authors such as Busa *et al.* (2001), Mendes & Chaves (2001), Vossen (2001), Marrafa (2002), O'Hara & Wiebe (2003), Veale (2003), Pederson & Sorensen (2006), among others.

The levels and elements of representation proposed here for inclusion in lexical entries will be used in the description of Portuguese verbs of movement, based on the wordnet developd for this class of verbs, and considering the established semantic elements that are part of their meaning, discussed in previous chapters.

However, and given the prominence of the relation of prepositions with verbs of this class, typically introducing SOURCE, GOAL, DIRECTON and PATH denoting arguments, further analyses of this POS are in order, and will be subject of the next chapter.

5. On prepositions

5.0 Introduction

In the previous chapters we focused on the description of verbs, in what concerns the lexicalization of movement concepts in Portuguese and the encoding of informational structures at the lexical entry level.

Capturing the base concept lexicalized by a given verb and the specific meaning content that it denotes with regard to its hyperonym contributes to determining the meaning components that are added to the base concept and allows us to identify the semantic elements that differentiate the nodes in a lexical-conceptual net, and the lexical semantic properties that may account for the semantic and syntactic behavior of the verbs at stake.

The levels and elements of representation and the generative mechanisms of GL provide an adequate, systematic and coherent modeling of the semantic content of verbs as well as of some of their basic syntactic properties. Also, the organization of Portuguese verbs of movement in a wordnet contemplates lexical inheritance, allowing the sharing of pertinent semantic and syntactic information within the net.

However, given the relevance of verbal argument structure, and specifically of the subcategorization properties of verbs and the type of constructions in which verbs of movement typically occur, it is necessary to account for arguments realized by prepositional phrases, since often the verbs in this class often select SOURCE, PATH, GOAL and DIRECTION denoting arguments, which are typically realized by prepositional phrases.

The analysis of prepositions has largely considered the relation between prepositions and the nouns they co-occur with (*on Thursday, in the morning*) or the verbs that select them (*dream of*, *care about*) (Veerspoor 1997), their own semantic description often being somewhat overlooked. Even so, many prepositions display a constant semantic content, which is crucial for the determination of PP meaning (*since February* vs. *until February*; *at home* vs. *from home*) (Bannard & Baldwin 2003).

Research on prepositions has taken three main directions, all concerning the semantic description of prepositions, and all considering their syntactic behavior in a more or less deeper level: large-scale symbolic accounts of preposition semantics (such as Dorr's (1997) 497 senses of English transitive and intransitive prepositions formalized in a lexical conceptual semantics framework, Canesson & Saint-Dizier's (2002) description of French prepositions in PrepNet, or Jensen & Nilsson's (2003) description of prepositions through a finite set of universal binary role relations); PP disambiguation accounts (such as O'Hara & Wiebe (2003)b account of PP tokens according to case-roles or McShane *et al.*'s (2005) ontological semantic analyzer for disambiguating homonym prepositions); and distributional accounts of preposition semantics (such as Bannard & Baldwin's (2003) work on particles and transitive prepositions for a valence-conditioned classification of English prepositions).

Profiting from already developed work on prepositions, this chapter is dedicated to the modeling of prepositions in the lexicon, regarding their integration in wordnets, and their semantic representation at lexical level, within the GL model. The first section of this chpater is dedicated to the integration of prepositions in WordNet.PT, based on previous research on ontological models for the representation of prepositions. Section 5.2 presents our treatment of prepositions in GL, based on Pustejovsky's (1995) introductory description of this POS lexical semantics and on current work on decompositional and semantic description of prepositions, making also use of the hierarchical relations established in the wordnet. The last section concerns the treatment of argument-marking prepositions[1] in wordnet considering also the informational structures of these items in GL.

5.1 Prepositions in wordnets

The first issue to consider when it comes to integrating prepositions in a wordnet is that this POS was not considered in the original WordNet 1.5. As stated before, WordNet evolved from the need for a comprehensive computational database which contemplated both word forms and meanings (see Miller 1998: xvi), and focused first on the open classes of nouns and verbs, having then integrating the modifier classes of adjectives and adverbs. This strategy left prepositions out.

Several researchers have, meanwhile, concentrated in prepositions and in the ontological organization of prepositions, adopting a similar approach to that of WordNet in the sense that prepositions are described according to their conceptual properties (see McShane *et al.* 2005, Saint-Dizier 2005, Jensen & Nilsson 2003, Canesson & Saint-Dizier 2002, for instance).

[1] Sag & Wasow (1999: 157).

5.1.1 Concepts denoted by prepositions

PrepNet (Saint-Dizier 2005, 2008) is such an example. PrepNet is a database for prepositions structured in two levels: the abstract notion level (conceptual level, language independent) and the language realization level (which deals with the realizations for various languages). Abstract notions are characterized by a name and an informal gloss, represented by Lexical Conceptual Structure (LCS) primitives viewed as linguistic macros and subdivided according to the several senses of each abstract notion. Abstract notions, on their turn, are organized in a first level that characterizes the semantic family of the notions (*localization*, *manner*, *quantity*, *company*, etc.), a second level that accounts for the different facets of each semantic family (*source*, *destination*, or *via*, for instance), and a third level that captures the modalities of a given facet (such as *basic manner*, *manner by comparison*, *manner with a reference point*, etc.). The language representation level includes syntactic frames (the syntactic subcategorization frames of the head of the prepositional phrase) and semantic and domain restrictions (selection restrictions for each argument in the frame).

PrepNet approach to the representation of the meaning of prepositions can be used as the base for integrating prepositions in wordnets, since the first level of representation considered in PrepNet (abstract notion) can help in the establishment of prepositional higher nodes in wordnet and the second level of description, in which prepositions are characterized with regard to the more abstract notion, can be used as a guideline to establish sets of hyponyms.

Naturally, the research and modeling of prepositions only makes sense in an integrated study of the lexicon, which further motivates the integration of prepositions in lexical models such as wordnets: "Although the study of prepositions is an interesting topic in itself, it is of much interest to investigate how this work can be integrated into larger frameworks such as FrameNet or VerbNet, and this is one of our major prospective." (Saint-Dizier 2008: 768).

From traditional grammars (see Cunha & Cintra (1984)) to state of the art models (Saint-Dizier 2008), prepositions are basically described as items that relate other items of a sentence. Jensen & Nilsson's (2003), for instance, propose a finite set of universal binary role relations to describe the semantic content of prepositions. In their perspective, prepositions denote a relation between the concept denoted by a given lexical item and semantic roles considered in a given ontology.

The following examples list the relations denoted by prepositions in a Portuguese traditional grammar (in (1)), the set of semantic roles assumed in Jensen & Nilsson (2003) (in (2)) and the abstract notions considered in PrepNet (in (3)).

(1) Traditional Grammar description of the meaning of prepositions (inspired in Cunha & Cintra (1984:554))

- Preposition Meaning
  - movement
    - spatial relation; ex: *Eu vou **para** Roma*. (I'm going **to** Rome)
    - temporal relation; ex: **De** tempos **a** tempos ele vinha a casa. (**From** time **to** time he came home)
    - notion relation; ex.: *O João chorou **de** raiva*. (John cried **in** anger)
  - situation
    - spatial relation; ex.: *O gato vive **no** sofá*. (The cat lives **in** the couch.)
    - temporal relation; ex.: *Tudo aconteceu **em** 24 horas*. (Everything happened **in** 24 hours)
    - notion relation; ex.: *A proposta foi recebida **com** medo*. (The offer was received **with** fear)

(2) Top ontoly of role relations and examples presented in Jensen & Nilsson (2003: 8)

AGENT	animate being acting intentionally (ex: *treatment **by** physician*)
CAUSE	inanimate force/actor
CAUSED_BY	inversed CAUSE
PATIENT	affected entity/effected entity (ex: *treatment **of** children.*)
PART_OF	part of whole/member of (ex: *side **of** the head, cells **in** the eye, agent **from** the CIA*)
COMPRISE	inverse PART_OF; whole constitued of parts
BY MEANS OF	means to end/instrument (ex: *treatment **with** medicine*)
SOURCE	source, origin, point of departure (ex: *haemorrhege **from** the intestine*)
PURPOSE	purpose
LOCATION	place, position (ex: *inflammation **of** the eyes*)
TEMPORALITY	temporal anchoring, duration, inception, etc. (ex: ***for** two days, **from** last year*)
MATERIAL	material (ex: *cushion **of** leather*.)
CHARACTERIZE	property ascription (ex: *children **with** diabetes*)

(3) Abstract notions and facets denoted by prepositions (Saint-Dizier 2008: 764-765)

- Localization:
  - source
  - destination
  - via/passage
  - fixed position
- Quantity
  - numerical or referential quantity
  - frequency and iterativity
  - proportion or ratio
- Manner
  - manners and attitudes
  - means (instrument or abstract)
  - imitation, agreement or analogy
- Accompaniment
  - adjunction
  - simultaneity of events
  - inclusion
  - exclusion
- Choice and exchange
  - exchange
  - choice or alternative
  - substitution
- Causality
  - cause
  - goal or consequence
  - intention
  - purpose
- Opposition
- Ordering
  - priority
  - subordination
  - hierarchy
  - ranking
  - degree of importance
- Instrument
- Other groups
  - theme
  - in spite of
  - comparison

The sets illustrated above show that, in spite of the different theoretical frames and goals pursued, prepositions are transversally regarded as relation indicators, and concepts such as space, temporality, causality, and so on, are generally considered. These ontological analyses can provide us with the top concepts susceptible to be lexicalized by Portuguese prepositions and are of great use in what concerns their lexical-conceptual organization in a wordnet. Based on these studies, we propose the integration of prepositions in WordNet.PT, according to the lexical and conceptual relations previously characterized and according to new tests proposed here to this POS.

5.1.2 Integrating prepositions in WordNet.PT

Following a similar approach to the one used for building a wordnet for Portuguese verbs of movement, we started by collecting the prepositions commonly used in this semantic domain, like *de* (≅ from)*, a* (≅ to)*, até* (≅ until/to)*, para* (≅ to, in the direction of)*, por* (≅ through)*, em* (≅ in)*, sobre* (≅ over)*, entre* (≅ between)*,* etc., considering also multiword expressions such as *acima de* (≅ above), *atrás de* (≅ behind), *ao lado de* (≅ next to), *por baixo de* (≅ under), *em direcção a* (≅ in the direction of), and so on, since these fixed expressions behave like prepositions (see Cunha & Cintra (1984); and Baldwin *et al.* (2009) for a recent analysis). In this work, we will consider only prepositional multiword expressions that do not undergo inflection, internal modification or word order variation, i.e. "words with spaces" (Sag *et al.* 2002). The examples in (4) illustrate these properties:

(4) a. O João colocou o livro mesmo ao lado das jarras.
(John placed the book exactly at.the side of the vases.) (≅ next to)

b. *O João colocou o livro mesmo aos lados das jarras.
(John placed the book exactly at.the$_{plural}$ sides of the vase.)

c. *O João colocou o livro ao lado mesmo das jarras.
(John placed the book at.the exactly side of the vases)

d. *O João colocou o livro ao lado esquerdo das jarras.
(John placed the book at.the left side of the vases)

e. *O João colocou o livro mesmo do lado às jarras.
(John placed the book exactly of.the side at.the the vases)

Also, most of prepositional multiword expressions that integrate nouns are syntactically marked since these occur with no determiners and thus do not constitute a saturated NP (Baldwin *et al.* (2009:126)):

(5) a. *em frente a* and not *na frente a*
(*in front of* and not *in.the front of*)

b. *em lugar de* and not *no lugar de*
(*in place of* and not *in.the place of*) (≅ instead of)

c. *em direcção a* and not *na direcção a*
(*in direction to* and not *in.the direction to*) (≅ in the direction of)

d. *a respeito de* and not *ao respeito de*
(*with respect to* and not *with.the respect to*)

Finally, prepositional multiword expressions can often be replaced by simple prepositions (as illustrated in (6)).

(6) a. O rato correu <u>em direcção a/para</u> a mesa.
(The mouse ran in the direction of/to the table)

b. O rato está <u>debaixo de/sob</u> a mesa.
(The mouse is under the table)

c. O rato está <u>em cima de/sobre</u> a mesa.
(The mouse is on top of the table)

As expected, the organization of prepositions in wordnets will be structured by synonymy relations, reflected in the members of each synset, and hyperonymy relations, reflected in the hierarchy levels, defined and exemplified as follows:

(7) a. P_1 is synonym of P_2 in C iff
if P_1 then P_2 and if P_2 the P_1

b. Se o livro está *sobre* a mesa então o livro está *em cima d*a mesa e se o livro está *em cima d*a mesa então o livro está *sobre* a mesa.
(if the book is *above* the table then the book is *on top of* the table and if the book is *on top of* the table then the book is *above* the table

{sobre, em cima de}$_{Prep}$ (≅ above, on top of)

c. Se ele falou *sobre* animais então ele falou *acerca de* animais e se ele falou *acerca de* animais então ele falou *sobre* animais.
(if he talked *about* animals then he talked *about* animals and if he talked *about* animals then he talked *about* animals)

{sobre, acerca de}$_{Prep}$ (≅ about)

(8) a. P_2 is hyponym of P_1 iff

i) P_2 is P_1+NP_i/$AdjP_j$/$AdvP_k$/PP_l

but

ii) P_1 is not P_2+NP_i/$AdjP_j$/$AdvP_k$/PP_l

b. *entre* é *em*+o espaço que separa objectos mas *em* não é *entre*+o espaço que separa objectos

(*between* is *in*+the space that separates objects, but *in* is not *between*+the space that separates objects)

{entre}$_{Prep}$ (≅ between) IS HYPONYM OF {em}$_{Prep}$ (≅ in)

c. Os animais estão entre as casas.

(The animals are between the houses)

c'. Os animais estão no espaço que separa as casas.

(The animals are in the space that separates the houses)

d. Os animais estão nas casas.

(The animals are in the houses)

d.' #Os animais estão entre o espaço que separa as casas.

(The animals are in the space that separates the houses)

Synonymy relations between prepositions are not very productive, even considering the synonymy notion bound to a given context. This is probably related to the fact that prepositions are a closed-class and seems to be conversely proportional to the highly polysemic behavior of prepositional expressions.

Prepositional synsets can also be related to each other by antonymy.

(9) a. P_1 is antonym of P_2 iff

i) P_1 and P_2 are co-hyponyms and

ii) P_1+NP_i/VP_j is the opposite of P_2+NP_i/VP_j and P_2+NP_i/VP_j is the opposite of P_1+NP_i/VP_j

ii) if P_1+NP_i/VP_j then not P_2+NP_i/VP_j and if P_2+NP_i/VP_j then not P_1+NP_i/VP_j

b. *em cima de* (≅ on top of) and *debaixo de* (≅ under) are both hyponyms of *em* (≅ in) and:

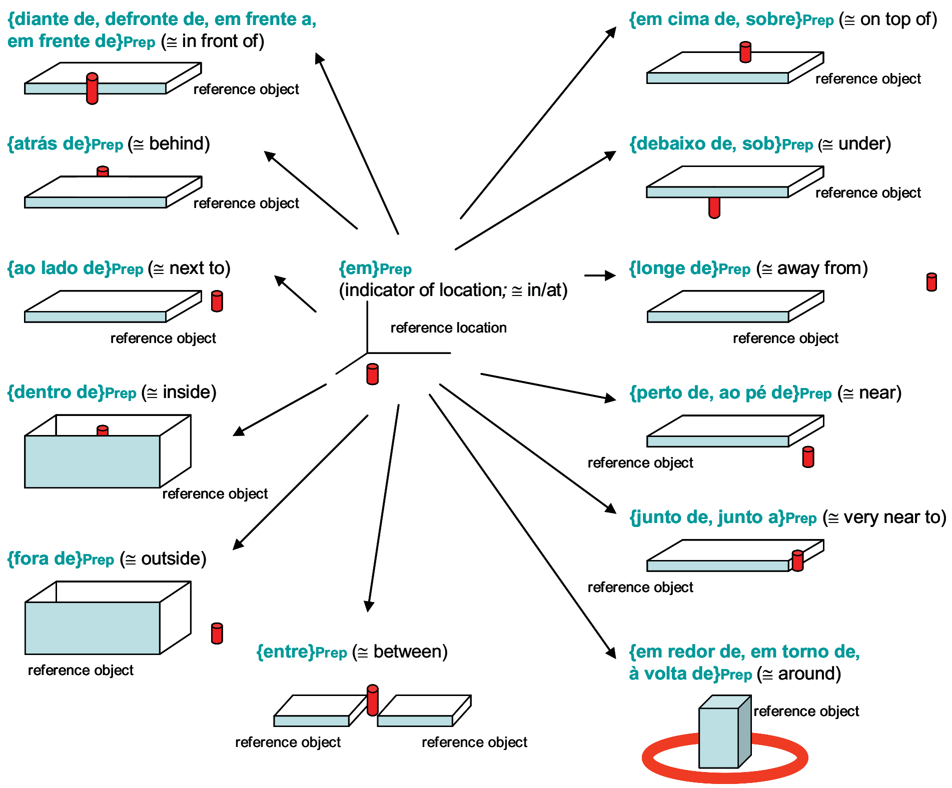
i) *em cima d*a mesa is the opposite of *debaixo d*a mesa and *debaixo d*a mesa is the opposite of *em cima d*a mesa.

ii) se o gato está *em cima d*a mesa então o gato não está *debaixo d*a mesa e se o gato está *debaixo d*a mesa então o gato não está *em cima d*a mesa.

(if the cat is *on top of* the table then the cat is not *under* the table and if the cat is *under* the table then the cat is not *on top of* the table) *em cima de* (≅ on top of) IS ANTONYM OF *debaixo de* (≅ under)

According to the relations available in the net, we modeled some subsets of prepositions whose denoted concepts are directly related to movement events. The first subnet concerns indicators of location, in (9).

(10) Hyponymy network of prepositional synsets denoting indicators of location



The Portuguese preposition *em* is the top node for this subnet, roughly corresponding to the English prepositions *in*/*at*. This preposition denotes the more general and underspecified concept indicator of location, which is then specified by its hyponyms.

The first observation to be made is that it is not easy to gloss these concepts without resorting to the lexical items we intend to describe[2]: if *entre* (≅ between) can be more or less artificially described as "*em* (in/at) the space that separates objects", prepositional expressions such as *debaixo de* (≅ under), *em cima de* (≅ on top of), *ao lado de* (≅ next to), *atrás de* (≅ behind), etc., are not as easily glossed:

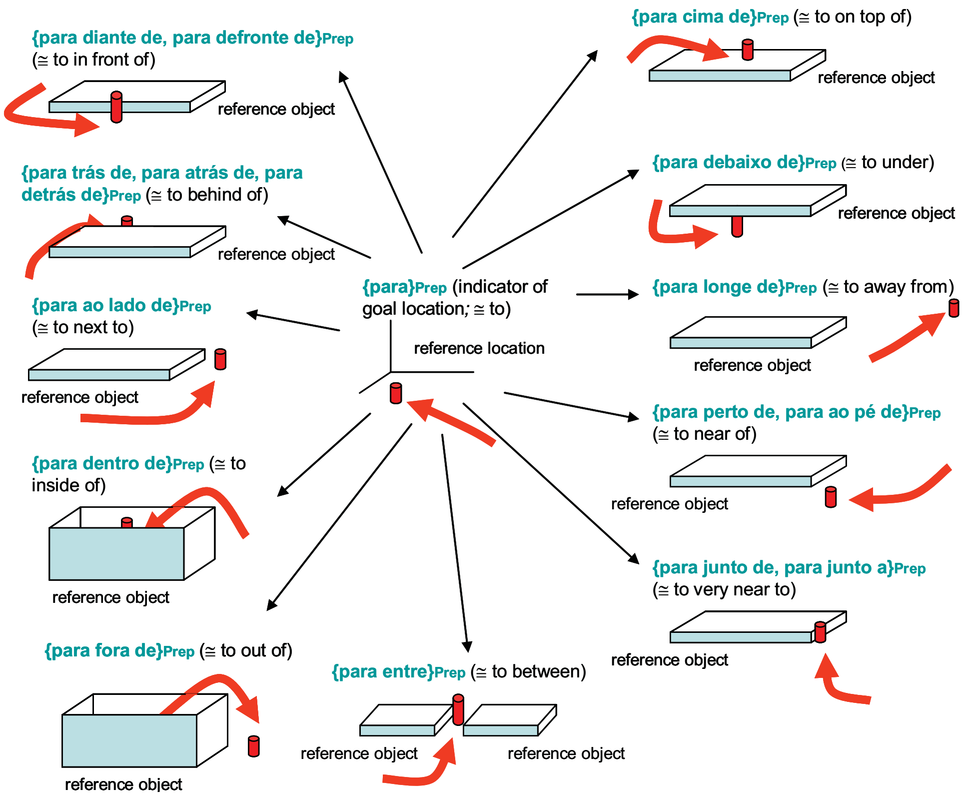
[2] Several authors use spatial models to describe the meaning of prepositions. See Galton (1993, 1997), Herzog (1995), Asher & Sablayrolles (1996), Lockwood *et al.* (2005), for instance.

(11) {*debaixo de, sob*}: *em* (in/at) the space that is under a given object?
em (in/at) the space that is below a given object?

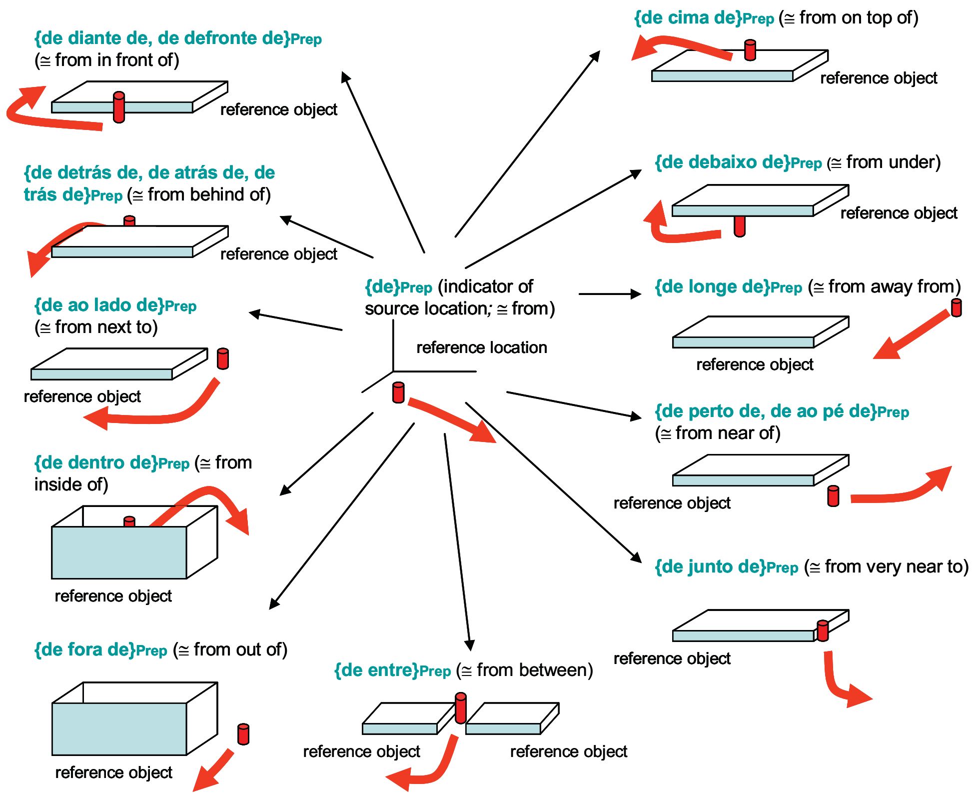
Given that the informal definition of the concepts denoted by the nodes in a wordnet consists basically in an user-friendly aid – the meaning of a node in the net emerges from its relative position in the network, i.e., from the lexical-conceptual relations established with the other nodes –, and that the lexical entries of prepositional expressions will also comprehend structured information in GL format, for the time being and for purposes of explanation we will only present the probable correspondences in English.

Also related to the indication of location, but in different subnets, we have prepositions denoting goal indicators and prepositions denoting source indicators, organized as follows (in (12) and (13) respectively):

(12) Hyponymy network of prepositional synsets denoting indicators of goal locations



(13) Hyponymy network for prepositional synsets denoting indicators of source locations



Although seeming quite similar to prepositions denoting indicators of location, prepositional expressions denoting goal and source locations do not result from the combination of the prepositions *para* (≅ to) and *de* (≅ from) with prepositions denoting location indicators (in (9)), since many combinations do not occur. For instance, it is not possible to express a source or goal location using the prepositions *de* or *para* + *em* (the top nodes of the three subtrees presented):

(14) a. *O João foi de em a escola para em a rua.
(John went from in the school to in the street)

A closer view also reveals that several combinations of elements fom the expressions in the subnets presented above are not possible:

(15) a. *para/de em cima de (≅ to/from on top of)

b. *para/de em baixo de (≅ to/from on under of)

c. *para/de em frente a (≅ to/from in front of)

d. para trás de/*de trás de/*em trás de (≅ to behind/from behind/in behind)

e. para debaixo de/de debaixo de/em *debaixo de (≅ to under/from under/under)

f. em torno de/*para torno de/*de torno de (≅ in around of/to around of/from around of)

Also, prepositions denoting indicators of source and goal locations are not hyponyms of $\{em\}_{Prep}$ (≅ in/at; location indicator):

(16) a. **para* é *em*+o local final mas *em* não é *para*+o local final
(*to* is *in*+the final location, but *in* is not *to*+the final location)

b. Os animais correram para as casas.
(The animals ran to the houses)

b'. *Os animais correram no local final as casas.
(The animals ran in.the final location the houses)

d. Os animais correram nas casas.
(The animals ran in the houses)

d.' *Os animais correram para o local final as casas.
(The animals ran to the final location the houses)

(17) a. *de é *em*+o local inicial mas *em* não é *de*+o local inicial
(*from* is *in*+the initial location, but *in* is not *from*+the initial location)

b. Os animais correram das casas.
(The animals ran from the houses)

b'. *Os animais correram no local inicial as casas.
(The animals ran in.the initial location the houses)

c. Os animais correram nas casas.
(The animals ran in the houses)

c.' *Os animais correram do local inicial as casas.
(The animals ran from the initial location the houses)

However, intuitively, the concepts of location, source location and goal location seem to be strongly related. This is the case given that moving to a final location (goal) causes being in that location, and, on the contrary, moving from a given location (source) causes not being in that location. Being so, it is possible to link these concepts in wordnets through cause relations.

The establishment of cause relations between prepositions requires a slight adjustment of the testing formulae, since prepositions on their own do not have reference potential and so, it is necessary to consider the preposition complement. In (18) below, we present a new formula for testing cause relation and the test application to the synsets $\{para\}_{Prep}$ (≅ to; indicator of goal) and $\{de\}_{Prep}$ (≅ from; indicator of source) and $\{em\}_{Prep}$ (≅ in/at; indicator of location) in (19).

(18) Test for CAUSE relation

P_1+N_i causes/has as consequence P_2+N_i, but not the converse.

(19) a) *para a rua* causes/has as consequence *na rua* but *na rua* does not cause/have as consequence *para a rua*. (*to the street* causes/has as consequence *in.the street* but *in.the street* does not cause/have as consequence *to the street*)

$\{para\}_{Prep}$ (≅ to)	CAUSES	$\{em\}_{Prep}$ (≅ in/at)	and
$\{em\}_{Prep}$ (≅ in/at)	IS CAUSED BY	$\{para\}_{Prep}$ (≅ to)	(non-factive)

b) *da rua* causes/has as consequence not-*na rua* but *na rua* does not cause/have as consequence not-*da rua*. (*from the street* causes/has as consequence not-*in.the street* but *in.the street* does not cause/have as consequence not-*from.the street*)

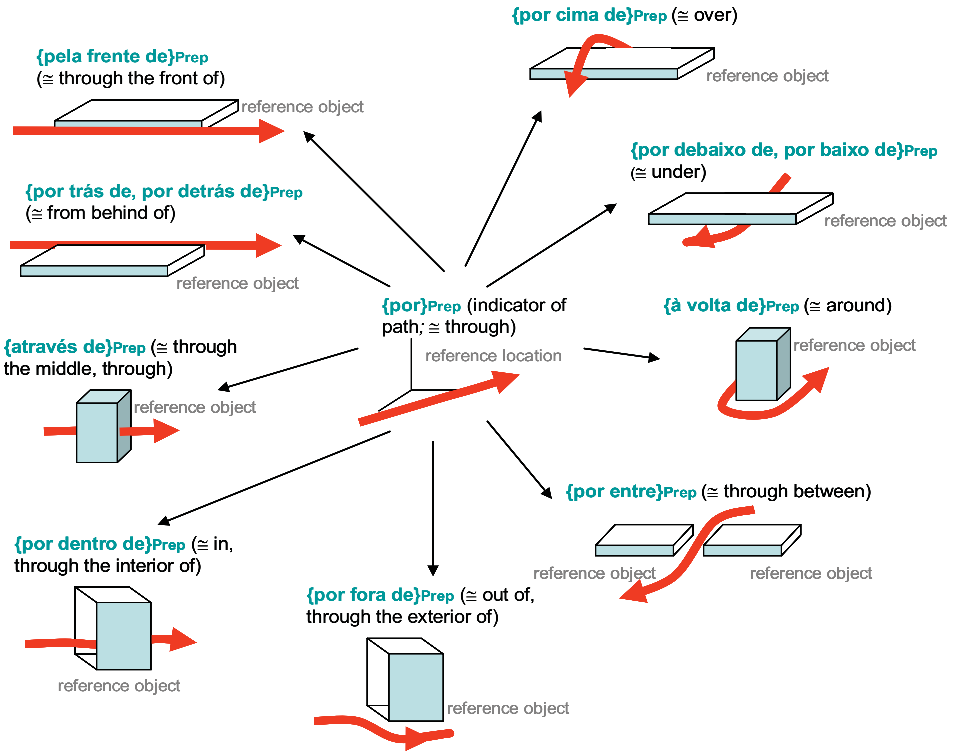
$\{de\}_{Prep}$ (≅ from)	CAUSES	$\{em\}_{Prep}$ (≅ in/at)	(negative) and
$\{em\}_{Prep}$ (≅ in/at)	IS CAUSED BY	$\{de\}_{Prep}$ (≅ from)	(negative) (non-factive)

Note that in order to test the cause relation between a prepositional synset indicator of source location and a prepositional synset indicator of location, in (19)b, it is also necessary to include negation, since the consequent state of moving from a given location amounts to not being in that location.

The *negation* label, as presented in the chapter 2, is used to explicitly express that a given relation does not hold. Typically, it is used to block unwanted implications, as non-inherited relations (Vossen 2002:16). The case presented here does not exactly correspond to the same situation, given that there is no prototypical relation to be inherited. Also, the relation established between $\{de\}_{Prep}$ (≅ from) and $\{em\}_{Prep}$ (≅ in/at) is that the first concept causes the negation of the last concept and not that the relation between the concepts does not hold. For this reason, it is only possible to express this relation indirectly, linking $\{de\}_{Prep}$ (≅ from) and $\{para\}_{Prep}$ (≅ to) as antonyms.

In what concerns concepts related to movement, it is also necessary to integrate the prepositional expressions that denote indicators of path and indicators of direction in the net. The following examples depict the organization of these prepositional items.

(20) Hyponym network for prepositions denoting indicator of path



(21) Prepositions denoting indicator of direction

{para 2, em direcção a}Prep (indicator of direction; ≅ *in the direction of*]

The set of Portuguese prepositions used in the domain of movement presented above is not exhaustive but represents common prepositional expressions used in this semantic field. The subnets of prepositions denoting indicators of location, source location, goal location and path are fairly comparable in size, whereas the prepositional expressions used for conveying the direction of movement considered here are only two. Note that directions, in Portuguese, as in other languages, are also often expressed by adverbial expressions such as *para cima* (≅ up, upwards), *para baixo* (≅ down, downwards), *para a frente* (≅ forward), *para trás* (≅ back, backwards), etc.

The same seems to be true in what concerns expressions denoting initial or final position co-occurring with change of position verbs: more often than not, initial or final positions are specified through adverbial and adjectival phrases rather than by prepositional ones:

(22) a. O João pôs o livro de pé/ao alto/deitado/na horizontal/de pernas para o ar/…
(John placed the book standing/up/laying/horizontally/upside down/…)

b. O João pôs-se de pé/de cócoras/de joelhos/de lado/…
(John placed himself standing/up/crouching/on his knees/on his side/…)

Another group of prepositional expressions that co-occur with Portuguese verbs of movement is composed of prepositions that can be described as quantity or duration denoting prepositions, which establish the length or duration of the movement event. In Portuguese, the common prepositional expressions that denote indicators of beginning and indicators of end are the ones presented in (23) below, and occur with location or time denoting noun phrases[3], as illustrated in (24).

(23) Prepositions denoting indicator of beginning and end

{de 2, desde}Prep
(indicator of beginning; ≅ from, since)

{até a}Prep
(indicator of end; ≅ to, until)

ANTONYM OF

(24) a. O João arrastou a caixa de/desde a casa até ao escritório.
(John dragged the box from the house to the office)

b. O João arrastou a caixa de/desde as 10 horas até às 11 horas.
(John dragged the box from/since 10 am to/until 11 am)

Typically, these expressions co-occur simultaneously in the same sentence, given that the measuring of the event requires the definition of both a starting and an ending point of the event. However, this is not always the case, since there are other contextual factors that may provide beginning or ending points of the event, without it being overtly expressed at syntax:

(25) a. O João arrastou a caixa desde/da praça/desde as 10h.
(John dragged the box from the square/since 10h ≅ John has been dragging the box from the square/ since 10h *until now*)

b. O João arrastou a caixa até à praça/às 10h.
(John dragged the box to the square/until 10h ≅ John dragged the box *since he started* until reaching the square/until 10h)

[3] The use of space denoting expressions in the temporal measuring of events has been traditionally considered (Cunha & Cintra 1983:553) and it is a phenomenon transversal to several languages. For more insights on this subject see, for instance, Radden (2004).

Regardless of other considerations concerning co-occurrence restrictions involving the verbs in study, to which we will return in chapters 6 and 8, the prepositions presented here are the most prominent subset of prepositions that occur with Portuguese verbs of movement and are directly related to the selection properties of these verbs. The integration of prepositions in wordnet allows us to directly relate them with the verbs in study, achieving a more complete description of the subcategorization properties of these verbs.

Besides, one of the advantages of relational models of the lexicon is the percolation of information through the net. Since wordnets can function as GL lexical inheritance structure, as discussed in the previous chapter, it is possible to automatically predict which prepositional expressions can introduce location, source, goal, etc., arguments, since a given argument can be realized by the indicated node or by any of its hyponyms:

(26)	O João pôs o livro	*em* a mesa.
	(John put the book	*in* the table)
		*debaixo d*a mesa(≅ *under* the table)
		entre a mesa(≅ *between* the table)
		*diante d*a mesa(≅ *in front of* the table)
		*atrás d*a mesa (≅ *behind* the table)
		*dentro d*a mesa (≅ *inside* the table)
		…

Note, however, that the specific meaning of hyponym prepositions naturally imposes constraints on prepositional complements. For instance, the preposition *entre* (≅ between) requires a complement that refers to plural sets of entities, since it indicates a location comprised between two or more entities, and so the well-formed sentence corresponding to the one presented above should be *John put the book between the tables.*

This way, it is possible to determine that the verb *pôr* (≅ put) selects an argument introduced by a preposition denoting an indicator of source location, the specific realization of the argument being conditioned by the semantic properties of the elements the preposition relates, corresponding in this case to the object denoted by the direct object of the verb, *the book* in the sentence in (26), and the object denoted by the complement of the preposition, in this case *the table.* This explains why sentences such as *John put the book inside the table* may be odd, or at least require the assumption that the table in question has an interior compartment, whereas sentences such as *John put the book inside the closet* may seem slightly redundant, as opposed to *John put the book in the closet*, since the container sense of the object denoted by *closet* constitutes one of its defining semantic properties.

This phenomenon illustrates the compositional aspect of language, reinforcing the need for informational structures at lexical entry level.

5.2 Prepositions in GL

Pustejovsky (1991, 1995) gives little attention to prepositions. His first reference to prepositions and their representation in the lexicon concerns a case of co-composition of directional prepositions and manner of motion verbs. In order to account for the transitional reading of the process denoting verb *float* in the sentence *the bottle floated into the cave*, the author, assuming that directional PPs act as a function over the verb, proposes the following representation of the PP *into the cave*:

(27)

$$\begin{bmatrix} \text{into the cave} \\ \text{ARGSTR} = \begin{bmatrix} \text{ARG}_1 = \boxed{1}\,[\text{physobj}] \\ \text{ARG}_2 = \boxed{2}\,[\text{the\_cave}] \end{bmatrix} \\ \text{EVENTSTR} = \begin{bmatrix} \text{E}_1 = \text{e}_1\text{: process} \\ \text{E}_2 = \text{e}_2\text{: state} \\ \text{RESTR} = <\alpha \\ \text{HEAD} = \text{e}_2 \end{bmatrix} \\ \text{QUALIA} = \begin{bmatrix} \text{FORMAL} = \text{at}(\text{e}_2, \boxed{1}, \boxed{2}) \\ \text{AGENTIVE} = \text{move}(\text{e}_1, \boxed{1}) \end{bmatrix} \end{bmatrix}$$

Pustejovsky (1995: 126)

This representation allows the author to derive the transitional reading of the sentence *The bottle floated into the cave* through co-composition (as discussed in the previous chapter, section 4.1.2.1). However, the semantic content of this phrase is not sufficiently motivated, namely with regard to the reasons for considering the meaning of this preposition equivalent to any transition denoting verb. Assuming this representation would allow us to predict that the sequence *the bottle into the cave* is the well-formed syntactic projection of this PP and that this sequence means *the bottle moved into the cave*.

As discussed in previous sections of this chapter, and as can be deducted from the AVM presented by Pustejovsky (1995: 126) to represent the prepositional phrase *into the cave*, prepositions establish a relation between two objects. For this reason, and as noticed by Saint-Dizier (2005, 2008), the description of prepositions must take into account elements not completely headed by the preposition, as the verb and its external argument (in the AVM in (27), ARG_1), which together with the complement of the preposition (ARG_2 in the AVM above) constitute the semantic objects that enter the relation denoted by the preposition. However, it does not seem reasonable to assume a predicate representation for prepositions, at least in what concerns goal and source denoting prepositions.

Saint-Dizier (2008) solves this issue by associating complex syntactic frames to prepositions, as shown in (28):

(28) Representation for *par, via* (≅ through) synset

Given X and Y, restricted respectively to concrete entities and location, and Verb, restricted to inherently directed motion:

```
VIA
'An entity X moving via a location Y'
representation: X : via(loc, Y)
[X(np,subj), Verb, Y(np,obj,optional), preposition, Z(np,obj2)]
```

(taken from Saint-Dizier(2008:765-766)

In GL, the relation denoted by the preposition is represented at qualia structure level, since this is the level that provides the semantic objects that define the meaning of a given lexical item, through semantic predicates that state the relation between the predicative term and the logical arguments of a given lexical item (stated at argument structure level), resulting in the semantic objects that define the meaning of a lexical item. Considering that the preposition para (≅ to) does not denote an event (it cannot replace state denoting items such as adjectives, process or transition denoting items such as verbs, nor event denoting nouns, see (29)), it does not have an event structure.

(29) a. O João está cansado. (John is tired).
a'. *O João está para a rua. (John is to the street)

b. O João está a correr/construir a casa. (John is running/building the house).
b'. *O João está a para a rua. (John is to the street)

c. A destruição da cidade aconteceu ontem. (The destruction of the city happened yesterday)
c'. *Para a rua aconteceu ontem. (To the street happened yesterday)

However, semantically full prepositions (as opposed to argument-marking prepositions, discussed ahead) introduce predicates that establish a relation between the argument of the preposition and an external argument (see Sag & Wasow (1999: 157, 184)). This way, the possible lexical entry for the preposition *para* (≅ to) can be exemplified as follows, being *r* introduced as the variable for *relation* (as *e* conventionally stands for event):

(30) para (indicator of goal; ≅ *to*)

$$\begin{bmatrix} \text{ARGSTR} = \begin{bmatrix} \text{P-ARG1} = \text{r} \\ \text{ARG1} = \text{y} = \begin{bmatrix} \text{objecto (object)} \\ \ldots \end{bmatrix} \end{bmatrix} \\ \\ \text{QUALIA} = \begin{bmatrix} \text{FORMAL} = \text{to(r, x, y), at(e}_1\text{, x, y)} \end{bmatrix} \end{bmatrix}$$

The semantic predicates stated as values of the Formal role represent the relation established by the preposition, *r*, and the objects it relates, *x* and *y*, (*to(r,x,y)*), but also the resulting event caused by this relation, namely *at(e_1, x, y).* Given that only one argument is stated at argument structure level, ARG_1 = y, x has to be externally instantiated, i.e., it is contextually defined. The AVM in (31) exemplifies this case.

(31) correr para a escola (run to school)

$$\begin{bmatrix} \text{ARGSTR} = \begin{bmatrix} \text{P-ARG1} = \text{e} : [\text{mover-se (move oneself)}] \\ \text{ARG1} = \boxed{1}\ \text{x} = [\text{entidade animada (animated entity)}] \\ \text{ARG2} = \text{z} = \begin{bmatrix} \text{para (to)} \\ \text{ARGSTR} = \begin{bmatrix} \text{P-ARG1} = \text{r} \\ \text{ARG1} = \text{y} = [\text{escola (school)}] \end{bmatrix} \\ \text{QUALIA} = [\text{FORMAL} = \text{to}(\text{r}, \boxed{1}, \text{y}), \boxed{2}\, \text{at}(\text{e}_1, \boxed{1}, \text{y})] \end{bmatrix} \end{bmatrix} \\ \\ \text{EVENTSTR} = \begin{bmatrix} \text{E1} = \text{e1} : \text{process} \\ \text{E2} = \boxed{2} \\ \text{HEAD} = \text{e1} \end{bmatrix} \\ \\ \text{QUALIA} = \begin{bmatrix} \text{AGENTIVE} = \text{correr}(\text{e1}, \text{x} : \text{object}, \text{z}) \\ \text{FORMAL} = \boxed{2} \end{bmatrix} \end{bmatrix}$$

This way, and also using co-composition, it is possible to derive the transitional meaning of the VP *run to school*, respecting, at the same time the semantic and syntactic properties of non event denoting prepositions such as *para* (≅ to). Note, nonetheless, that the GL model is a model of the lexicon that does not comprehend syntactic processes. Thus, the phrasal representations are only used here for purposes of explanation and should not be regarded as making any claims regarding syntactic derivations.

Interestingly, the argument of the preposition, in our example, is quite underspecified, since it is not possible to further define its semantic properties. It seems that, although the PP refers a goal location, the preposition does not select a noun denoting location. In other words, the *location* meaning is also achieved compositionally.

(32) O rato correu para o armário.
(the mouse ran to the closet)

As expected, prepositions denoting indicators of source locations, such as *de* (≅ from), have similar lexical entries but denoting the opposite concept, which amounts to the caused event being the exact opposite of the one caused by the relation denoted by *para* (≅ to):

(33) de (indicator of source; $\cong$ *from*)

$$\begin{bmatrix} \text{ARGSTR} = \begin{bmatrix} \text{P-ARG}_1 = \text{r} \\ \text{ARG}_1 = \text{y} = \begin{bmatrix} \text{objecto (object)} \\ \ldots \end{bmatrix} \end{bmatrix} \\ \\ \text{QUALIA} = \left[\text{FORMAL} = \text{from(r, x : object, y)}, \neg\text{at(e}_1\text{, x, y)}\right] \\ \end{bmatrix}$$

At this point, and returning to the lexical entries of the prepositions listed above, it is necessary to observe that prepositions denoting indicators of source and goal locations, on the one hand, and prepositions denoting indicators of direction and path, on the other, select different kinds of arguments. That is, while prepositions denoting indicators of goal and source locations establish a relation between a given location and an object, prepositions denoting indicators of direction and path establish a relation between an event argument, more specifically a process, and a specific direction or path. The AVMs in (34) and (35) present our proposal for the lexical entries of the prepositions *por* ($\cong$ through) and *para/em direcção a* ($\cong$ to/in the direction of).

(34) por (indicator of path; $\cong$ *through*)

$$\begin{bmatrix} \text{ARGSTR} = \begin{bmatrix} \text{P-ARG1} = \text{r} \\ \\ \text{ARG1} = \text{z} = [\text{via } (\textit{path})] \vee [\text{local } (\textit{location})] \end{bmatrix} \\ \\ \text{QUALIA} = \left[\text{FORMAL} = \text{through(r, e : process, z)}\right] \\ \end{bmatrix}$$

(35) para/em direcção a (indicator of direction; $\cong$ *to/in the direction of*)

$$\begin{bmatrix} \text{ARGSTR} = \begin{bmatrix} \text{P-ARG}_1 = \text{r} \\ \\ \text{ARG}_2 = \text{z} = [\text{direcção } (\textit{direction})] \vee [\text{objecto } (\textit{object})] \end{bmatrix} \\ \\ \text{QUALIA} = \left[\text{FORMAL} = \text{to(r, e : process, z)}\right] \\ \end{bmatrix}$$

Both paths and directions can be expressed by the relevant preposition plus nouns that lexicalize specific paths (*road*, *way*, etc.) and directions (*North*, *South*, *up*, and so on) or, compositionally, the arguments being in this case underspecified. In the case of prepositions denoting indicators of path, this possibility is represented by the disjunction of the possible top nodes *via* (path) and *local* (location) which constitutes the ARG_1 value, in (34). In the case of prepositions denoting indicators of direction this possibility is represented by the disjunction of the possible top nodes *direcção* (direction) and *objecto* (object; physical entity) in ARG_1, as illustrated in (35).

The prepositions discussed so far have stricter co-occurrence restrictions (they typically occur with change of location verbs) than prepositional expressions denoting indicators of location, beginning and end, which can occur with verbs from other semantic domains. Moreover, indicators of source and goal locations establish a relation between two entities (a physical object and other physical entities interpreted as locations) syntactically expressed by NPs; indicators of path, direction, beginning and end establish a relation between an event and paths (or locations interpreted as paths), directions (or physical entities interpreted as directions) and a given point in time or space, typically expressed by NPs; and indicators of location can establish a relation between two entities (a physical object and a location) or between an event and a location.

The prepositions denoting indicators of beginning and end, determine the relation between an event and a point in time or space, establishing the beginning or ending point of a given event. The AVMs in (36) and (37) illustrate our proposal for the lexical entries of these items. Note, that we chose to leave the arguments of these expressions underspecified given that they may correspond to particular lexicalizations of temporal or location concepts (*morning* or *Lisbon*), but they may also correspond to phrases (*11 o'clock* or *the end of the street*):

(36) de, desde (indicator of beginning; ≅ from, since)

$$\begin{bmatrix} \text{ARGSTR} = \begin{bmatrix} \text{P-ARG1} = \text{r} \\ \text{ARG1} = \text{z} = [\textit{time}] \vee [\textit{location}] \end{bmatrix} \\ \\ \text{QUALIA} = [\text{FORMAL} = \text{start\_point(r, e, z)}] \\ \end{bmatrix}$$

(37) até a, a (indicator of ending; ≅ until)

$$\begin{bmatrix} \text{ARGSTR} = \begin{bmatrix} \text{P-ARG1} = \text{r} \\ \\ \text{ARG1} = \text{z} = [\textit{time}] \vee [\textit{location}] \end{bmatrix} \\ \\ \text{QUALIA} = [\text{FORMAL} = \text{end\_point(r, e,z)}] \\ \end{bmatrix}$$

Also, although establishing the end or the starting point of a given event, prepositions denoting indicators of beginning and end do not necessarily cause a final state. That is, there is no final state that results from the relation denoted by these prepositions (see (38)), but the measuring of the duration of an event, as illustrated by the sentences in (39).

(38) a) #*até às 10h* cause/has as consequence *às 10h* but *às 10h* does not cause/have as consequence *até às 10h* (*until 10h* causes/has as consequence *at 10h* but *at 10h* does not cause/have as consequence *until 10h*)

b) #*desde as 10h* causes/has as consequence not-*às 10h* but not-*às 10h* does not cause/have as consequence *desde as 10h* (*since 10h* causes/has as consequence not-*at 10h* but not-*at 10h* does not cause/have as consequence *since 10h*)

(39) a) O João anda desde as 10h/Lisboa (John has/had been walking since 10h/Lisbon) $\rightarrow$ The walking event has the duration of the interval of time comprised between 10h/being in Lisbon and a reference point to be determined (if the verb is in the past tense) or unkown (if the verb is in the present tense).

b) O João andou até às 10h/Lisboa (John walked until 10h/Lisbon) $\rightarrow$ The walking event has the duration of the interval of time comprised between 10h/being in Lisbon and a given starting point of the event (typically determined contextually).

c) O João andou desde as 10h/Lisboa até às 12h/ao Porto (John walked since 10h/Lisbon until 12h/Oporto) $\rightarrow$ The walking event has the duration of the interval of time comprised between 10h/being in Lisbon and 12h/being in Oporto.

The determination of the starting or ending point of the event has also aspectual implications, reflected in the verb tense with which these expressions can occur. Sentences that express only the starting point of the event can occur with verbs in the present tense and refer to an ongoing event: *John has/had been walking since 10h* $\rightarrow$ *John is still walking*; sentences that express the ending point of the event occur with verbs in the past tense and refer to a non-ongoing event: *John walked until 10h* $\rightarrow$ *John is no longer walking*.

Finally, the preposition *em* ($\cong$ in/at), denoting an indicator of location, is a new case altogether, since it denotes a state. Prepositional phrases headed by this item can replace state denoting items such as adjectives, providing evidences for this assumption:

(40) a. O João está cansado. (John is tired).
a'. O João está na rua. (John is in the street)

For this reason, we propose the representation in (41) for this preposition.

(41) em (indicator of location; $\cong$ *in/at*)

$$\begin{bmatrix} \text{ARGSTR} = \begin{bmatrix} \text{P-ARG1} = \text{e} \\ \text{ARG1} = \text{z} = [\text{objecto } (\textit{object})] \end{bmatrix} \\ \\ \text{EVENTSTR} = \begin{bmatrix} \text{E1} = \text{e1} : \text{state} \\ \text{HEAD} = \text{e1} \end{bmatrix} \\ \\ \text{QUALIA} = [\text{FORMAL} = \text{at}(\text{e1}, \text{x} : \text{object} \vee \text{event}, \text{z})] \end{bmatrix}$$

As discussed earlier in this section, in this case the *location* meaning is also achieved compositionally, the preposition requiring only that its argument refers to a physical entity (represented here by the node *object* (objecto). Also, and given that prepositions denoting indicators of location can relate events and physical entities to a given location, for purposes of explanation, the external semantic object, x, is represented here as corresponding to an object or to an event, through disjunction.

The representation of prepositions in GL and their integration in a wordnet allows us to account for verbal arguments introduced by prepositions. The prepositions described above can be stated as values of true arguments of some verbs (*tirar x de+local* ≅ take x from+source location, *pôr x em+local* ≅ put x in+source location, etc.), or as default-arguments of change of location verbs (*Ele empurrou a cadeira para a parede* ≅ he pushed the chair to the wall), as it will be further discussed in the next chapter, always maintaining their meaning, a fact which supports their integration in a relational lexicon. Following the approach in Verspoor (1997), we assume that the semantic contribution of prepositional phrases is consistent across uses, regardless of their status as complements or adjuncts. Verspoor (1997) concerns mainly dative prepositional phrases, divided in three types of prepositional datives: complement PPs, pseudo-complement PPs, and adjunct PPs. The semantic content of all three types of PPs is constant, the type of prepositional modification being determined via lexical rules.

Extending this analysis to the prepositions presented here, it is possible to account for prepositional complements, stated as values of true arguments in verbal lexical entries, prepositional adjuncts being determined by general rules (syntactic and semantic) that regulate co-occurrence restrictions of lexical items, and thus not reflected in the lexical semantic and syntactic properties of verbs.

Given that our interest in prepositions is necessarily related to the representation of verbal items in the lexicon, there is also a subset of prepositions to be considered, namely the so-called argument-marking prepositions.

5.3 Argument-marking prepositions

One of the main reasons leading to the little attention dedicated by grammatical tradition to prepositions when it comes to their semantic content is directly related to the semantically empty or argument-marking prepositions, i.e., prepositions whose only function is to mediate between a given predicate and its arguments (Sag & Wasow 1999: 157):

(42) a. O rapaz gostou de cães.
the boy liked PREP dogs

b. O rapaz sonhou com cães.
the boy dreamt of dogs

c. O rapaz aproximou-se dos cães.
the boy came closer to the dogs.

The sentences in (42) illustrate this case, namely cases in which the presence of the preposition is language dependent (in (42)a); cases in which preposition choice does not correspond to the typical equivalent in other languages (in (42)b, where the Portuguese preposition *com* corresponds to the English preposition *of*, instead its English translation *with*); and cases where the argument-marking preposition is homonym of the preposition denoting the opposite semantic content (in (42)c, where the preposition *de* marks a goal location denoting argument, whereas the semantically full preposition *de* denotes an indicator of source location.

Being idiosyncratic, i.e. language dependent and not permutable by any other preposition, argument-marking prepositions are said to form a semantic component with the verb, since it is the verb+preposition that attributes case to the selected NP (or DP) (see Neeleman (1997)). This proposal results in complex lexical entries for verbs such as *gostar de* (≅ like), *sonhar com* (≅ dream of) and *aproximar-se de* (≅ go closer), for instance.

However, and as underlined by Godoy (2008), at syntactic level, these prepositions form constituents with the selected NP, and not with the verb, as shown by the tests below:

(43) a. De cães, o rapaz gosta. (≅ PREP dogs, the boy likes)

b. Com cães, o rapaz sonhou. (≅ Of dogs, the boy dreamt)

c. Dos cães, o rapaz aproximou-se. (≅ To the dogs, the boy moved closer)

(44) a. O rapaz gosta de cães e ela também gosta. (≅ the boy likes PREP dogs and she also likes)

b. O rapaz sonhou com cães e ela também sonhou. (≅The boy dreamt of dogs and she also dreamt)

c. O rapaz aproximou-se dos cães e ela também se aproximou. (≅The boy moved closer to the dogs and she also moved closer)

(45) a. O rapaz gosta de cães e de gatos. (≅ the boy likes PREP dogs and PREP cats)

b. O rapaz sonhou com cães e com gatos. (≅The boy dreamt of dogs and of cats)

c. O rapaz aproximou-se dos cães e dos gatos. (≅The boy moved closer to the dogs and to the cats)

These examples show that, although compulsory in the syntactic realization of a given verb, the argument-marking prepositions do not form semantic and/or syntactic components with the verb that subcategorize for them: on the one hand, having no semantic content these prepositions do not contribute to the semantic content denoted by the VP; on the other, they form syntactic constituents with the NP and not with the verb. Following Godoy's (2008) approach, we will consider that these prepositions are not visible at semantic level, existing solely at syntactic level.

For these reasons, and in what concerns the representation of argument-marking prepositions at lexical entry level, we will consider these prepositions to be semantically empty lexical items, directly related to verbs that select them as values for true arguments. This way, it is possible to account for their mandatory syntactic realization, their syntactic omission being licensed only when recoverable by context, as well as for the tight syntactic relation these prepositions establish with the NP.

(46) aproximar-se

$$
\begin{bmatrix}
\text{ARGSTR} = \begin{bmatrix}
\text{P-ARG}_1 = \text{e}_1 : \boxed{1} \begin{bmatrix} \text{mover-se } (\textit{move}) \\ \ldots. \end{bmatrix} \\
\text{ARG}_1 = \text{x} = [\text{animated entity}] \\
\mathbf{ARG_2} = \mathbf{y} = \begin{bmatrix} \mathbf{de} \\ \mathbf{ARGSTR} = [\mathbf{ARG_1} = \mathbf{x} = [\mathbf{physical\ entity}]] \end{bmatrix}
\end{bmatrix} \\
\\
\text{EVENTSTR} = \begin{bmatrix}
\text{E}_1 = \text{e}_1 : \boxed{1} \\
\text{E}_2 = \text{e}_2 : \text{state} \\
\text{RESTR} = < \alpha \\
\text{HEAD} = \text{e}_1
\end{bmatrix} \\
\text{QUALIA} = \begin{bmatrix}
\text{FORMAL} = \text{aproximar-se\_state}(\text{e}_2, \text{x}, \text{y}) \\
\text{AGENTIVE} = \text{aproximar-se}(\boxed{1}, \text{x})
\end{bmatrix}
\end{bmatrix}
$$

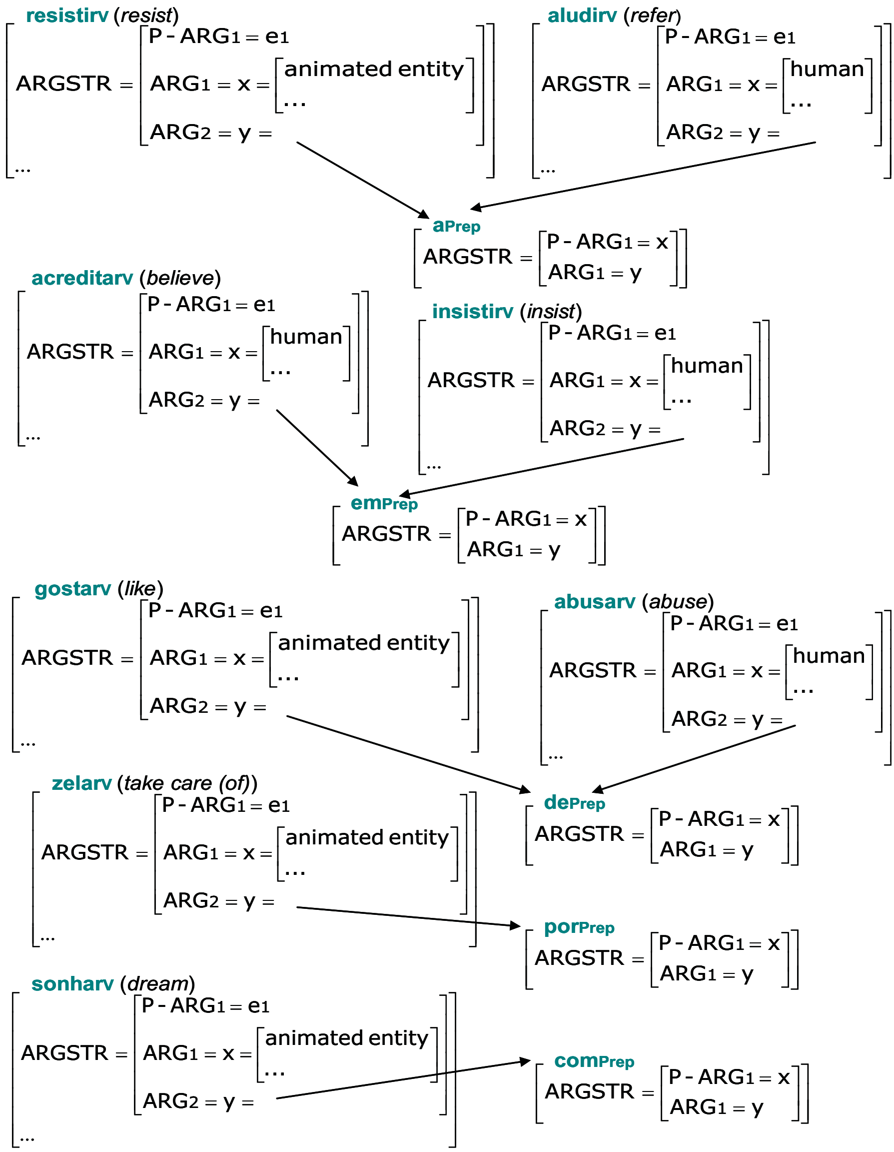
An approach such as this raises the issue of the representation of these lexical items in lexical-conceptual relational lexica, since the prepositions at stake do not denote any concept. However, their inclusion in the lexicon can be motivated by two main reasons of different order:

First, as idiosyncratic items, these prepositions are acquired by children in a similar process as all other lexical items, since their distribution and/or meaning does not result from the regular application of the rules available in natural languages (see Godoy (2008)).

Second, argument-marking prepositions constitute a small and closed set of items, necessarily connected to the predicates that require their syntactic realization. This way, and in spite of being represented as autonomous entries in the lexicon, the collection and treatment of argument-marking prepositions is always related to the collection and treatment of other POS, typically verbs.

For these two reasons, we propose to include these items in a wordnet as part of the set of prepositional items, but with extremely underspecified lexical entries. Although not linked to other lexical items through hyperonymy or synonymy relations, these prepositions are related to other lexical items in the net through argument structure relations (as proposed in chapter 7):

(47) Example of relations linking argument-marking prepositions in a wordnet



5.4 Conclusions

Benifiting from work already developed on prepositions, in this chapter we propose the modeling of prepositions in wordnets and their semantic representation at lexical entry level, within the GL model. The integration of prepositions in wordnets follows previous research on ontological models for the representation of prepositional meaning, and the concepts denoted by prepositions are those consensually considered in traditional grammars and state of the art models.

Focusing on prepositions that commonly co-occur with Portuguese verbs of movement, we follow the assumption that prepositions establish a relation between entities, independently of them denoting events or not, and propose their modeling using the levels of representation at our disposal. In a coherent and unified manner, we account simultaneously for semantically full prepositions that can introduce verbal arguments or adjuncts, since the first are determined in verbal lexical entries, and thus directly connected to the verbal nodes in the net. Also, and given their relevance for the description of verbal properties, we propose to integrate argument-marking prepositions in wordnets in order to achieve a richer representation of the selection properties of the verbs in study.

However, and given the goals of the present work, the treatment of prepositions discussed here is not by far exhaustive or comprehensive. Our main intention, which we consider accomplished, is to motivate the integration of prepositions in wordnets and to provide lexical descriptions for these items in order to support our treatment of verb selection properties.

6. Portuguese verbs of movement in GL

6.0 Introduction

Following the analysis presented in the previous chapters of this dissertation, in this chapter we focus on the modeling of Portuguese verbs of movement in GL. This modeling reflects, in various degrees, the observations perceived so far, namely, that:

i) organizing verbs in a wordnet establishes a lexical-conceptual motivated inheritance structure that allows for the direct determination of what is specific to a given item and for the observation of the patterns that may emerge from the lexicalization or specification of semantic elements;

ii) the informational structures at the lexical entry level are able to reflect both semantic and syntactic properties of the represented lexical items, accounting for divergent behaviors of verbs of the same class;

iii) the modeling of lexical items of a given POS is not independent from that of the others of different POS with which they occur.

The semantic and syntactic information stated in the lexical entries of verbs, according to the formulations of the GL discussed in chapter 4 and given the lexical inheritance structure established through hyperonymy, obviates what is inherited from the hyperonym and what is particular to the hyponym, i.e. , in what way does the hyponym specifies the concept denoted by the hyperonym. This analysis concerns lexical inheritance and divergent semantic and syntactic properties, namely Aktionsart properties, differences in argument structure and different qualia structure values. The semantic and syntactic properties of the verbs in study are directly related to the types of arguments (in GL terms), which has effects on the relations established between a given lexical item and others in the lexicon.

For these reasons, the present chapter is divided in three main sections. The first section addresses the general cases, demonstrating lexical inheritance and percolation of information in a hierarchical lexicon. The second section focuses on the lexicalization of the semantic components determined earlier, the incorporation of semantic restrictions on these elements and how these properties are represented in the lexical structures. Section 3 is dedicated to the formalization of verbal co-hyponyms compatibility through qualia unification, establishing the differences between nominal and verbal qualia unification. And, finally, section 4 presents the conclusions of this task.

6.1 Lexical inheritance through hyponymy

In this section we present the lexical entries for Portuguese verbs of movement, implementing the proposals presented so far, namely lexical inheritance structure through the indication of the hyperonym as value of the proper argument in the argument structure of lexical items and the direct relation of the arguments of a given lexical item to available structures (other AVMs) in the lexicon, including prepositional structures. This way, we model the meaning specificities of the hyponyms in GL structures, considering the semantic components determined.

The lexical entries of the verbs in study must state both the general semantic and syntactic properties of these verbs, contemplating at the same time the differences emerging from the specification of the concepts denoted by the higher nodes in the net.

As established in chapter 2, the set of Portuguese verbs of movement has two distinct top nodes, corresponding to the more underspecified lexicalizations of the concepts of change of location and of change of position. These two nodes determine, thus, the base lexical semantic properties that will be inherited by all their respective subordinates.

Let us start by the top node for change of locations verbs, lexicalized by the verbs *mover, deslocar* (see chapter 2, section 2.2.1). These verbs denote a process (see (1)),

(1) a. O homem moveu/deslocou o objecto durante 10 minutos.
(The man moved the object for 10 minutes)

b. O homem está (agora) a mover/deslocar o objecto → o homem (já) moveu/deslocou o objecto.
(The man is (now) moving the object→ the man has (already) moved the object)

and select two true arguments, an underspecified first argument corresponding to any entity and a second argument denoting a physical entity, as exemplified in (2)a and b and (2)c and d, respectively:

(2) a. *Moveu/deslocou o objecto.
(Moved the object)

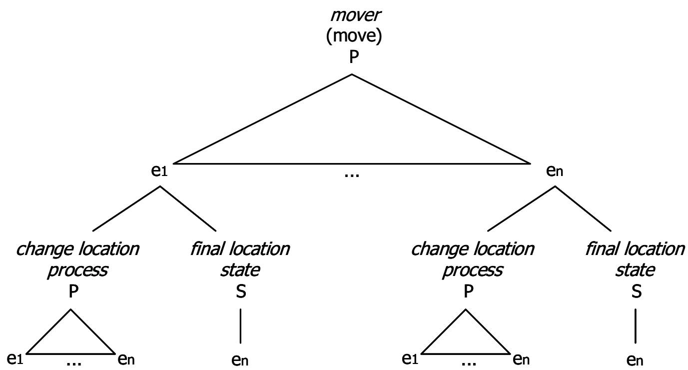
b. *O homem moveu/deslocou.
(The man moved)

c. O homem/o vento/o calor moveu/deslocou o objecto.
(The man/the wind/the heat moved the object)

d. O homem moveu/deslocou o objecto/#o vento/#a vontade.
(The man moved the object/the wind/the will)

According to the definition presented in the chapter 1 of this work, verbs of change of location denote a movement event where the initial location in which the entity starts the movement event is different from the final location in which the entity is when the movement event ends, i.e., an event where an entity moves from location A to location B (≠ A). This definition is consistent with transition type events, since it comprehends a final state of being in a given location. For this reason, and given the Aktionsart properties of the verbs *deslocar* and *mover* tested here in (1), it seems that change of location verbs, in fact, denote complex processes which are composed of the repetition of transition type events:

(3)



This structure, as discussed in chapter 4, sections 4.1.1.2 and 4.1.1.3, is not directly reflected in the event structure level of representation since the events that constitute the transitions are not direct subevents of the process denoted by *mover*. However, and as the co-occurrence with expressions denoting SOURCE and GOAL demonstrates, the initial location in which an entity starts the movement and the final location in which an entity is at the end of the movement event (corresponding to the final state of the last transition subevent of the process), are directly related to change of location verbs.

(4) O homem moveu/deslocou a caixa ([da estrada]$_{SOURCE}$ [para o carro]$_{GOAL}$)
(The man moved the box ([from the road]$_{SOURCE}$ [to the car]$_{GOAL}$))

Given that the verbs *mover* and *deslocar* do not require the syntactic realization of these expressions to form well-formed sentences, GOAL and SOURCE denoting expressions can be represented as default arguments, since they correspond to parameters required by the logical expressions in the qualia structure of the lexical item that can be syntactically expressed by subtyping or specification processes. The AVM in (5) below presents our proposal for the top node of verbs denoting change of location.

(5) {mover, deslocar}$_V$ (≅ move)

$$\begin{bmatrix} \text{ARGSTR} = \begin{bmatrix} \text{P - ARG1} = \text{e} \\ \\ \text{ARG1} = \text{x} = [\{\text{entidade}\}_N \text{ (entity)}] \\ \text{ARG2} = \text{y} = [\{\text{objecto}\}_N \text{ (object)}] \\ \text{D - ARG1} = \text{u} = \begin{bmatrix} \{\text{de}\}_{Prep} \text{ (from)} \\ \text{ARGSTR} = [\text{ARG1} = \text{w} = [\{\text{entidade concreta}\}_N \text{ (physical entity)}]] \\ \ldots \end{bmatrix} \\ \text{D - ARG2} = \text{v} = \begin{bmatrix} \{\text{para}\}_{Prep} \text{ (to)} \\ \text{ARGSRT} = [\text{ARG1} = \text{z} = [\{\text{entidade concreta}\}_N \text{ (physical entity)}]] \\ \ldots \end{bmatrix} \end{bmatrix} \\ \text{EVENTSTR} = \begin{bmatrix} \text{E1} = \text{e1}: \text{process} \\ \text{HEAD} = \text{e1} \end{bmatrix} \\ \\ \text{QUALIA} = \begin{bmatrix} \text{AGENTIVE} = \text{change\_location}(\text{e1}, \text{x}, \text{y}) \\ \text{FORMAL} = \text{at}(\text{e2}, \text{y}, \text{w}), \text{at}(\text{e3}, \text{y}, \text{z}) \end{bmatrix} \end{bmatrix}$$

Note that the syntactic expression of source and goal default arguments is necessarily achieved through specification processes, since their realization requires the definition of their internal argument, *w* and *z* respectively. The syntactic expression of these arguments changes the aspectual value of the sentences in which the verbs occur, forcing transition readings.

The top node for change of position verbs, on its turn, is lexicalized by the verb *mexer* (see chapter 2, section 2.2.1). This verb also denotes a process (see (6)),

(6) a. O homem mexeu o objecto durante 10 minutos.
(The man moved the object for 10 minutes)

b. O homem está (agora) a mexer o objecto → o homem (já) mexeu o objecto.
(The man is (now) moving the object→ the man has (already) moved the object)

and also selects two true arguments, an underspecified first argument denoting any entity and a second argument denoting a physical entity:

(7) a. *Mexeu o objecto.
(Moved the object)

b. *O homem mexeu.
(The man moved)

c. O homem/o vento mexeu o objecto.
(The man/the wind moved the object)

d. O homem mexeu o objecto/#o vento/#a vontade.
(The man moved the object/#the wind/#the will)

Given that the change of position concept does not entail change of location, i.e., it concerns changing the location of an entity or parts of it within a reference location (see chapter 2, section 2.2.1), SOURCE and GOAL correspond to the notions of initial and final position. The resulting lexical entry is represented in the following AVM:

(8) {mexer}$_V$ (≅ move; change position)

$$\begin{bmatrix} \text{ARGSTR} = \begin{bmatrix} \text{P-ARG1} = \text{e} \\ \text{ARG1} = \text{x} = [\{\text{entidade}\}_\text{N}\,(\text{entity})] \\ \text{ARG2} = \text{y} = [\{\text{objecto}\}_\text{N}\,(\text{object})] \\ \text{D-ARG1} = \text{u} = [[\text{initial position}]] \\ \text{D-ARG2} = \text{v} = [[\text{final position}]] \end{bmatrix} \\ \text{EVENTSTR} = \begin{bmatrix} \text{E1} = \text{e1}:\text{process} \\ \text{HEAD} = \text{e1} \end{bmatrix} \\ \text{QUALIA} = \begin{bmatrix} \text{AGENTIVE} = \text{change\_position}(\text{e1}, \text{x}, \text{y}) \\ \text{FORMAL} = \text{in}(\text{e2}, \text{y}, \text{w}), \text{in}(\text{e3}, \text{y}, \text{z}) \end{bmatrix} \end{bmatrix}$$

Note that, in this case, the expression of the initial and final positions is not directly linked to a given node in the net since there seems to be no lexicalization of these concepts, or a specific lexical item that introduces them.

Being top nodes, the values of the proper arguments of these verbs do not establish any subtyping relation, that is, there is no hyperonym node to consider or from which these items inherit information. The true arguments are associated to specific nodes in the net, defining their semantic properties and syntactic expression. The arguments can be realized by the lexical expression to which they are associated with or by any of their hyponyms (as established in chapter 4, section 4.2).

Given the hierarchical organization of the lexicon through hyperonymy/hyponymy relations, it is expected that the hyponyms denote the meaning of their hyperonyms plus more. The lexical structure of the hyperonym is available to the hyponym through the hyponymy relation established at the proper argument level: '*x: y → x is hyponym of y*'.

The AVMs presented in (9) and (10) exemplify how the information is inherited. In the first case, the node *abanar* (≅ shake), hyponym of the node *mexer* (≅ change position), inherits the list of arguments of its hyperonym, shadowing the default arguments and adding to it a shadow argument corresponding to the lexicalization of MANNER (*abanar* is *mexer* how? With impulses in different directions).

(9) $\{abanar\}_V$ ($\cong$ shake)

$$
\left[\begin{array}{l}
\text{ARGSTR} = \left[\begin{array}{l}
\text{P-ARG}_1 = e : \left[\begin{array}{l}
\{\text{mexer}\}_V(\text{change position}) \\
\text{ARGSTR} = \left[\begin{array}{l}
\text{P-ARG}_1 = e \\
\text{ARG}_1 = x = \boxed{1}\left[\{\text{entidade}\}_N\,(\text{entity})\right] \\
\text{ARG}_2 = y = \boxed{2}\left[\{\text{objecto}\}_N\,(\text{object})\right] \\
\text{D-ARG1} = u = [[\text{initial position}]] \\
\text{D-ARG2} = v = [[\text{final position}]]
\end{array}\right] \\
\text{EVENTSTR} = \left[\begin{array}{l}
\text{E}_1 = \boxed{3}\, e_1 : \text{process} \\
\text{HEAD} = e_1
\end{array}\right] \\
\text{QUALIA} = \left[\begin{array}{l}
\text{AGENTIVE} = \text{change\_position}(e_1, x, y) \\
\text{FORMAL} = \boxed{4}\,\text{in}(e_2, y, u),\ \text{in}(e_3, y, v)
\end{array}\right]
\end{array}\right] \\
\text{ARG}_1 = \boxed{1} \\
\text{ARG}_2 = \boxed{2} \\
\text{S-ARG}_1 = u = \left[\begin{array}{l}
[\text{de uma posição}]\ (\text{from one position}) \\
\ldots
\end{array}\right] \\
\text{S-ARG}_2 = v = \left[\begin{array}{l}
[\text{para a posição oposta}]\ (\text{to the opposite position}) \\
\ldots
\end{array}\right] \\
\text{S-ARG}_3 = t = \left[[\text{with impulses in different directions}]\right]
\end{array}\right] \\
\text{EVENTSTR} = \left[\begin{array}{l}
\text{E}_1 = \boxed{3} \\
\text{HEAD} = \boxed{3}
\end{array}\right] \\
\text{QUALIA} = \left[\begin{array}{l}
\text{AGENTIVE} = \text{abanar}(\boxed{3}, \boxed{1}, \boxed{2}, u, v, t) \\
\text{FORMAL} = \boxed{4}
\end{array}\right]
\end{array}\right]
$$

In the AVM representation system, the sharing of information between two lexical entries is codified through numerical indexes that represent equal values. In this case, the true arguments are directly inherited from the hyperonym, 1 and 2, and so is the base process in the event structure, 3 , and the Formal role value, 4. The value of the Agentive role defines the meaning of the lexical items of this node and can be read as follows: *abanar* is *mexer* from one position to the opposite one with impulses in different directions. Recovering the definition of the hyperonym verb, it results in the paraphrase: *abanar* is one entity changing the position of an object from one position to the opposite one with impulses in different directions.

The option of not stating this semantic predicate as having a complex description such as *abanar(mexer(*e_1*, x, y),u, v, z)*, is directly related to the mapping function associated to the predicates in qualia structure. Assuming the principles defined in chapter 4, i.e. that arguments are listed from the less oblique – in subject syntactic position – to the more oblique one, according also to the type of arguments (proper arguments, true arguments, shadow arguments, default arguments) and to the lexical items that realize them (established through the connection to other structures in the lexicon) this description is not consistent with the realization of the verb *abanar*. This verb does not select a first true argument that corresponds to the projection of the lexical entry of a verb: it selects two true arguments, corresponding to the ones selected by its hyperonym, and incorporates shadow arguments that reflect its meaning specificity with regard to the event denoted by its hyperonym, from which it inherits the event structure (number and type of events). The node {balançar}$_V$ ($\cong$ sway), on its turn, inherits the true arguments from its hyperonym ({abanar}$_V$ ($\cong$ shake)), but also its event structure, namely the process type event.

(10) {balançar} ($\cong$ sway)

$$
\left[\begin{array}{l}
\text{ARGSTR} = \left[\begin{array}{l}
\text{P-ARG1} = \text{e}: \left[\begin{array}{l}
\{\text{abanar}\}_{\text{V}}\ (\text{shake}) \\
\text{ARGSTR} = \left[\begin{array}{l}
\text{P-ARG1} = \text{e}: [\{\text{mexer}\}_{\text{V}}\ (\text{move})] \\
\text{ARG1} = \text{x} = \boxed{1}\ [\{\text{entidade}\}_{\text{N}}\ (\text{entity})] \\
\text{ARG2} = \text{y} = \boxed{2}\ [\{\text{objecto}\}_{\text{N}}\ (\text{object})] \\
\text{S-ARG1} = [[\text{from an initial position}]] \\
\text{S-ARG2} = [[\text{to a final initial position}]] \\
\text{S-ARG3} = \text{z} = [[\text{with impulses in different directions}]]
\end{array}\right] \\
\text{EVENTSTR} = \left[\begin{array}{l}
\text{E1} = \boxed{3}\ \text{e1}: \text{process} \\
\text{HEAD} = \text{e1}
\end{array}\right] \\
\text{QUALIA} = \left[\begin{array}{l}
\text{AGENTIVE} = \text{abanar}(\text{e1}, \text{x}, \text{y}, \text{z}) \\
\text{FORMAL} = \boxed{4}\ \text{in}(\text{e2}, \text{y}, \text{u}), \text{in}(\text{e3}, \text{y}, \text{v})
\end{array}\right]
\end{array}\right] \\
\text{ARG1} = \boxed{1} \\
\text{ARG2} = \boxed{2} \\
\text{S-ARG1} = \text{w} = [\{\text{suavemente}\}_{\text{Adv}}\ (\text{smoothly})]
\end{array}\right] \\
\text{EVENTSTR} = \left[\begin{array}{l}
\text{E1} = \boxed{3} \\
\text{HEAD} = \boxed{3}
\end{array}\right] \\
\text{QUALIA} = \left[\begin{array}{l}
\text{AGENTIVE} = \text{balançar}(\boxed{3}, \boxed{1}, \boxed{2}, \text{w}) \\
\text{FORMAL} = \boxed{4}
\end{array}\right]
\end{array}\right]
$$

Note that shadow arguments correspond to semantic parameters lexicalized in the lexical item. For this reason, the hyponym node, inheriting the event arguments from its hyperonym, as well as the semantic predicates that defines them, necessarily inherits the meaning properties corresponding to the shadow arguments.

The recursive use of available lexical structures through hyponymy relation, corresponding here to the hyponym relation established at the proper argument level – which, on its turn is also related by hyponymy to a higher node – allows the percolation of information through the hyponymy trees, enabling a coherent and economic codification of the information.

Given this account of lexical inheritance, in the next section we will focus on the encoding of the lexicalization of semantic components discussed in chapter 3, that is, as exemplified above, the semantic content that distinguishes a hyponym from its hyperonym, as well as of the resulting syntactic and semantic properties.

6.2 Encoding the lexicalization of semantic components

6.2.1 Lexicalization and argument type

As discussed in chapter 3, semantic incorporation occurs in different degrees affecting the syntactic expression of predicates also at different degrees. The meaning specificity of the hyponym verb can express the lexicalization of a given component, resulting in the addition of a shadow argument in the argument structure of the hyponym verb (a semantic parameter fully incorporated in the semantic content of the lexical item); or it can express restrictions on the semantic components considered, resulting in the addition of a new true argument in the argument structure of the hyponym or in stricter semantic restrictions on a given argument, for instance. Let us consider the hyperonym tree presented in chapter 3, section 3.2.2, repeated here in (11), and the AVMs for the hyponym nodes {pôr, colocar}$_V$ (≅ put), {meter}$_V$ (≅ put inside of) and {enjaular}$_V$ (≅ cage), in (12), (13) and (15)[1].

(11) Fragment of the wordnet with information on verbs' overt arguments

{mover, deslocar}v [change location]
V(ARG$_1$, ARG$_2$)

- **{pôr, colocar}v** [*mover* to a given final location; ≃ *put*] **+GOAL**
  V(ARG$_1$, ARG$_2$, ARG$_3$)
  - **{meter}v** [*pôr* inside of; ≃ *put inside of*] **+DIRECTION**
    V(ARG$_1$, ARG$_2$, ARG$_3$)
    - **{enjaular}v** [*meter* in a cage; ≃ *cage*) **+GOAL**
      V(ARG$_1$, ARG$_2$)

[1] For space and clarity reasons, and to avoid large AVMs with repeated information, in this subsection we will leave underspecified the information of the top nodes, marking in bold the information that is inherited by the hyponym.

(12) {pôr, colocar}$_V$ (≅ put)

$$
\left[\begin{array}{ll}
\text{ARGSTR} = & \left[\begin{array}{l}
\text{P-ARG}_1 = \text{e} : [\{\text{mover, deslocar}\}_V \text{ (move)}] \\
\mathbf{ARG_1} = \mathbf{x} = [\mathbf{\{entidade\}_N\ (entity)}] \\
\mathbf{ARG_2} = \mathbf{y} = \boxed{1}[\mathbf{\{objecto\}_N\ (object)}] \\
\mathbf{D\text{-}ARG_1} = \mathbf{u} = \left[\begin{array}{l}
\mathbf{\{de\}_{Prep}\ (from)} \\
\mathbf{ARGSTR} = \mathbf{z} = [\mathbf{\{entidade\ concreta\}_N\ (physical\ entity)}]
\end{array}\right] \\
\text{S-ARG}_1 = \text{v} = \left[\begin{array}{l}
\mathbf{\{para\}_{Prep}\ (to)} \\
\mathbf{ARGSTR} = \left[\begin{array}{l}
\mathbf{P\text{-}ARG_1 = r} \\
\mathbf{ARG_1} = \boxed{\mathbf{2}}
\end{array}\right] \\
\text{QUALIA} = [\text{FORMAL} = \text{to(r, } \boxed{1}, \boxed{2}), \boxed{4})]
\end{array}\right] \\
\text{ARG}_3 = \text{w} = \left[\begin{array}{l}
\{\text{em}\}_{\text{Prep}} \text{ (in/at)} \\
\text{ARGSTR} = \left[\begin{array}{l}
\text{P-ARG}_1 = \text{e} \\
\text{ARG}_1 = \text{x} = \boxed{2}[\text{entidade concreta (physical entity)}]
\end{array}\right] \\
\text{EVENTSTR} = \left[\begin{array}{l}
\text{E}_1 = \boxed{3}\text{e}_1 : \text{state} \\
\text{HEAD} = \text{e}_1
\end{array}\right] \\
\text{QUALIA} = [\text{FORMAL} = \boxed{4}\text{at(e}_1, \boxed{1}, \boxed{2})]
\end{array}\right]
\end{array}\right] \\
\text{EVENTSTR} = & \left[\begin{array}{l}
\mathbf{E_1 = e_1 : process} \\
\text{E}_2 = \boxed{3} \\
\text{RESTR} = < \alpha \\
\text{HEAD} = \text{e}_1
\end{array}\right] \\
\text{QUALIA} = & \left[\begin{array}{l}
\text{AGENTIVE} = \text{pôr(e}_1, \text{x, y, v, w)} \\
\text{FORMAL} = \mathbf{at(e_3,}\ \boxed{\mathbf{1}}\mathbf{, z)}, \boxed{4}
\end{array}\right]
\end{array}\right]
$$

(13) {meter} (≅ put inside of)

$$
\left[\begin{array}{ll}
\text{ARGSTR} = & \left[\begin{array}{l}
\text{P-ARG}_1 = \text{e} : [\{\text{pôr, colocar}\}_V \text{ (put)}] \\
\mathbf{ARG_1} = \mathbf{x} = [\mathbf{\{entidade\}_N\ (entity)}] \\
\mathbf{ARG_2} = \mathbf{y} = \boxed{\mathbf{1}}[\mathbf{\{objecto\}_N\ (object)}] \\
\mathbf{D\text{-}ARG_1} = \mathbf{u} = \left[\begin{array}{l}
\mathbf{\{de\}_{Prep}\ (from)} \\
\mathbf{ARGSTR} = \mathbf{z} = [\mathbf{\{entidade\ concreta\}_N\ (physical\ entity)}]
\end{array}\right] \\
\text{S-ARG}_1 = \text{v} = \left[\begin{array}{l}
\{\text{para dentro de}\}_{\text{Prep}} \text{ (to inside)} \\
\text{ARGSTR} = \left[\begin{array}{l}
\text{P-ARG}_1 = \text{r} \\
\text{ARG}_1 = \boxed{2}
\end{array}\right] \\
\text{QUALIA} = [\text{FORMAL} = \text{to\_inside\_of(r, } \boxed{1}, \boxed{2}), \boxed{4})]
\end{array}\right] \\
\mathbf{ARG_3} = \mathbf{w} = \left[\begin{array}{l}
\mathbf{\{em\}_{Prep}\ (in/at)} \\
\mathbf{ARGSTR} = \left[\begin{array}{l}
\mathbf{P\text{-}ARG_1 = e} \\
\mathbf{ARG_1 = x} = \boxed{\mathbf{2}}[\mathbf{entidade\ concreta\ (physical\ entity)}]
\end{array}\right] \\
\mathbf{EVENTSTR} = \left[\begin{array}{l}
\mathbf{E_1} = \boxed{\mathbf{3}}\mathbf{e_1 : state} \\
\mathbf{HEAD = e_1}
\end{array}\right] \\
\mathbf{QUALIA} = [\mathbf{FORMAL} = \boxed{\mathbf{4}}\mathbf{at(e_1,}\ \boxed{\mathbf{1}}, \boxed{\mathbf{2}})]
\end{array}\right]
\end{array}\right] \\
\text{EVENTSTR} = & \left[\begin{array}{l}
\mathbf{E_1 = e_1 : process} \\
\text{E}_2 = \boxed{3} \\
\text{RESTR} = < \alpha \\
\text{HEAD} = \text{e}_1
\end{array}\right] \\
\text{QUALIA} = & \left[\begin{array}{l}
\text{AGENTIVE} = \text{meter(e}_1, \text{x, y, v, w)} \\
\mathbf{FORMAL} = \mathbf{at(e_3,}\ \boxed{1}\mathbf{, z)}, \boxed{\mathbf{4}}
\end{array}\right]
\end{array}\right]
$$

These nodes show different degrees of incorporation with different repercussions in their syntactic realization. The node {pôr, colocar}$_V$ (≅ put) incorporates restrictions on GOAL, requiring the realization of the specific location through a prepositional true argument introduced by the preposition *em* (≅ in/at) (indicator of location). This specification results also in the increase of the event arguments in the event structure, resulting in a transition type event: the verbs {pôr, colocar}$_V$ (≅ put) denote a transition type event, in which the final state subevent directly corresponds to the state denoted by the third true argument. Note, also, that the realization of the default argument in other change of location verbs, by defining the end point of the event, carries also aspectual shifts and results in accomplishments denoting sentences:

(14) a. O homem moveu/deslocou o objecto para a casa em 10 minutos.
(The man moved the object to the house in 10 minutes)

b. O homem está (agora) a mover/deslocar o objecto para a casa → o homem (ainda) não moveu/deslocou o objecto para a casa.
(The man is (now) moving the object para a casa → the man has not (yet) moved the object to the house)

The hyponym node {meter}$_V$ (≅ put inside of), on its turn, lexicalizes DIRECTION, reflected in the incorporation of a more specific goal indicator, but maintains the number and type of arguments of the hyperonym, i.e. it shows no visible repercussions at the syntactic realization level.

The lower hyponym node, {enjaular}$_V$ (≅ put inside of a cage), in (15) below, lexicalizes a specific goal location, a *cage*, resulting in the shadowing of its hyperonym third true argument. As it is possible to see in the AVM below, this corresponds only to the specification of the goal location, the argument of the prepositional phrase introduced by *em* (≅ in/at), which carries also the specification of the type of entity that can be *caged* – animals – through its Telic role value. The verb *enjaular* (≅ cage) has two true arguments instead of the three of its hyperonym.

(15) {enjaular} (≅ cage)

$$\left[\begin{array}{ll}
\text{ARGSTR} = & \left[\begin{array}{ll}
\multicolumn{2}{l}{\text{P - ARG}_1 = \text{e} : \left[\{\text{meter}\}_\text{V}\ \text{(put)}\right]} \\
\multicolumn{2}{l}{\textbf{ARG}_\mathbf{1} = \mathbf{x} = \left[\textbf{\{entidade\}}_\mathbf{N}\ \textbf{(entity)}\right]} \\
\multicolumn{2}{l}{\textbf{ARG}_\mathbf{2} = \mathbf{y} = \boxed{1}\left[\{\text{animal}\}_\text{N}\ \text{(animal)}\right]} \\
\textbf{D - ARG}_\mathbf{1} = \mathbf{u} = & \left[\begin{array}{l}
\textbf{\{de\}}_\mathbf{Prep}\ \textbf{(from)} \\
\textbf{ARGSTR} = \mathbf{z} = \left[\textbf{\{entidade concreta\}}_\mathbf{N}\ \textbf{(physical entity)}\right]
\end{array}\right] \\
\text{S - ARG}_1 = \text{w} = & \left[\begin{array}{ll}
\multicolumn{2}{l}{\textbf{\{em\}}_\mathbf{Prep}\ \textbf{(in/at)}} \\
\textbf{ARGSTR} = & \left[\begin{array}{ll}
\multicolumn{2}{l}{\textbf{P - ARG}_\mathbf{1} = \mathbf{e}} \\
\textbf{ARG}_\mathbf{1} = \mathbf{y} = & \left[\begin{array}{l}
\{\text{jaula}\}\text{N (cage)} \\
\text{ARGSTR} = [\text{P - ARG1} = \text{x} : \{\text{objecto}\}\text{N (object)}] \\
\text{QUALIA} = [\text{TELIC} = \text{imprison(e1, x, y : animal)}]
\end{array}\right]
\end{array}\right] \\
\textbf{EVENTSTR} = & \left[\begin{array}{l}
\mathbf{E}_\mathbf{1} = \boxed{2}\ \mathbf{e}_\mathbf{1} : \textbf{state} \\
\textbf{HEAD} = \mathbf{e}_\mathbf{1}
\end{array}\right] \\
\multicolumn{2}{l}{\textbf{QUALIA} = \left[\textbf{FORMAL} = \boxed{3}\ \textbf{at}(\mathbf{e}_\mathbf{1}, \boxed{1}, \mathbf{y})\right]}
\end{array}\right]
\end{array}\right] \\
\text{EVENTSTR} = & \left[\begin{array}{l}
\mathbf{E}_\mathbf{1} = \mathbf{e}_\mathbf{1} : \textbf{process} \\
\mathbf{E}_\mathbf{2} = \boxed{3} \\
\textbf{RESTR} = <\boldsymbol{\alpha} \\
\textbf{HEAD} = \mathbf{e}_\mathbf{1}
\end{array}\right] \\
\text{QUALIA} = & \left[\begin{array}{l}
\text{AGENTIVE} = \text{enjaular}(\text{e}_1, \text{x}, \text{y}, \text{w}) \\
\textbf{FORMAL} = \textbf{at}(\mathbf{e}_\mathbf{3}, \boxed{1}, \mathbf{z}), \boxed{3}
\end{array}\right]
\end{array}\right]$$

Besides illustrating different degrees of lexicalization and the consequent syntactic differences, these verbs also show a specific pattern of syntactic realization of their shadow arguments.

By definition, shadow arguments can be syntactically expressed only by subtyping or specification processes. Typically, in the case of shadow arguments corresponding to concepts denoted by nouns, these processes allow the syntactic realization of the specific denoting noun (even if mediated by a preposition) as long as it is modified, with no implications on the remainder of the arguments realization (see (16)). The same is true for adverbial shadow arguments (in (17)). In fact, when it comes to MANNER denoting adverbial shadow arguments, in many cases the modification by the same adverbial modifier is allowed, resulting in a new modification of the MANNER concept already denoted by the verb.

(16) O rapaz enjaulou o leão numa jaula **dourada**.
(The boy caged the lion in a **golden** cage)

(17) O rapaz balançou o berço **suavemente/ muito suavemente**.
(The boy swayed the crib **smoothly/very smoothly**)

However, this is not the case for the verbs *pôr, colocar* and *meter*, represented above. The syntactic expression of the shadow arguments is not compatible with the simultaneous expression of the third true argument, indicator of location:

(18) *O rapaz pôs/colocou a caixa [**para dentro d**a casa]$_{S\text{-}ARG1}$ [**n**a casa]$_{ARG3}$.
(The boy put the box [**into** the house]$_{S\text{-}ARG1}$ [**in** the house]$_{ARG3}$.)

(19) *O rapaz meteu o livro [**muito para dentro d**a gaveta] $_{S\text{-}ARG1}$ [**n**a gaveta]$_{ARG3}$.
(The boy put the book [**much to the interior of** the drawer]$_{S\text{-}ARG1}$ [**in** the drawer]$_{ARG3}$.)

This seems to emerge from the fact that the indication of goal location results from the combination of the shadow argument (indicator of goal location) with the true argument (indicator of location), which necessarily share the same argument, since the shadow argument corresponds to part of the meaning of the verbs themselves: the concept denoted by these verbs requires the indication of a given final location, but does not establish which specific one (contrary to what occurs in the case of the verb *enjaular* (≅ cage), for instance), and requires the realization of an indicator of location. Also, it must be taken into account that both shadow and true arguments are prepositional phrases referring to locations, and that the indication of goal location entails the indication of a location. Therefore, the realization of the shadow argument of these verbs must consider their compatibility with the true argument realization. That is, if realized, through specification processes – for instance the realization of any of the hyponyms of the indicator of goal location that constitutes the shadow arguments of these verbs – the shadow argument provides the final location, making the realization of the true argument redundant:

(20)	O rapaz pôs/colocou a caixa	**para dentro de**	a casa.
	(*The boy put the box*	*to the interior of/ inside of*	*the house*.)
		para debaixo de (≅ *to under*)	
		para cima de (≅ *to on top of*)	
		para diante de (≅ *to in front of*)	
		...	

Note, nonetheless, that this is not the usual case. Typically, when there is further specification of the final location, this is achieved through the specification of the location, i.e. the true argument introduced by the preposition denoting indicator of location, typically through hyponym expressions:

(21) a. Eu próprio colocarei **debaixo d**os teus pés o tapete mais precioso,…
(I myself will put **under** your feet the most precious carpet, …)

b. [Ele] e o seu empregado (…) colocaram dois blocos de mármore (…) **junto a**o restaurante…
([He] and his employee (…) put two marbol blocks (…) **near** the restaurant…)

c. … aqueles empilhando, estes colocando cinco tijolos **sobre** cada ombro dos companheiros.
(… those making piles, these puting five bricks **over** each partner's shoulders)

d. Tivemos até proprietários (…) de restaurantes que pediram para colocar baias **em frente a**os seus estabelecimentos, porque os carros impediam a entrada…
(We even had (…) restaurant owners that ask us to put barriers **in front of** their establishments, because the cars blocked the entrance…)

e. A rega subsuperficial, em que a água é aplicada através de emissores colocados **abaixo d**a superfície do solo…
(The sub superficial watering, in which the water is applied through emissors put **under** the ground surface…)

f. Por causa do campo magnético, não coloque a televisão **perto de** colunas de som e de aquecedores de ferro ou aço.
(Due to the magnetic field, do not put the television set **near** loud speakers and steel or iron heaters.)

(22) a. [Ele] meteu a carta **entre** muitos outros papéis;…
([He] put the letter **between** lots of other papers; …)

b. … a minha opinião é que teu pai, se trouxe o dinheiro, não o tem em casa. Meteu-o **debaixo de** alguma fraga aí da serra …
(… my opinion is that your father, if he brought the Money, he doesn't keep it at home. He put it **under** some rock there in the hills…)

(examples from *CRPC – Reference Corpus of Contemporary Portuguese*, CLUL)

After discussing these examples, which illustrate some particular aspects of the lexicalization of the semantic components in the verbs in study, we will focus on the individual semantic components, on the regularities found in Portuguese verbs of movement, and on how these are mirrored in the lexical entries of the items analyzed.

6.2.2 Semantic and syntactic properties

In this section we focus on showing how the determination of the relevant semantic information contributes to the explanation and prediction of the syntactic behavior of Portuguese verbs of movement.

We will start by addressing the cases that reflect the lexicalization of the semantic elements directly related to movement, namely GOAL, SOURCE, DIRECTION and PATH, continuing then to the other semantic elements – common to verbs of other semantic domains –, MANNER, FIGURE, GROUND, INTENTION and CAUSE, providing examples and discussing, whenever pertinent, the patterns encountered.

6.2.2.1 GOAL and SOURCE

As described in chapter 3 of this dissertation, and as illustrated by some of the lexical entries presented in the previous section, the specification of the semantic components GOAL and SOURCE result, or occur, in transition denoting verbs. The specification of GOAL and SOURCE – whether by lexicalization, represented by shadow arguments; whether by the incorporation of semantic restrictions on these elements, expressed by the increase of the number of true arguments – implies the specification of a limit to the denoted event: the initial or final location/position, corresponding to the final state of the movement event.

Given that the top nodes of change of location verbs and of change of position verbs denote process type events, the specification of GOAL and SOURCE results in an Aktionsart shift. The nodes {pendurar-se}$_V$ (≅ move into a suspended position), {deitar-se}$_V$ (≅ move into a horizontal position), {estender-se}$_V$ (≅ move into a straight, unfolded position) are some examples of change of position verbs which, by lexicalizing the notion of GOAL, have Aktionsart properties different from those of their hyperonyms.

It what concerns the event structure of these verbs, the different event type denoted is also reflected in the number of subevents that constitute the denoted transition.

(23) {tirar} (≅ take)

$$
\left[\begin{array}{ll}
\text{ARGSTR} = & \left[\begin{array}{l}
\text{P-ARG}_1 = e: [\{\text{mover, deslocar}\}_V \text{ (move)}] \\
\text{ARG}_1 = x = [\{\text{entidade}\}_N \text{ (entity)}] \\
\text{ARG}_2 = y = \boxed{1} [\{\text{objecto}\}_N \text{ (object)}] \\
\text{ARG}_3 = u = \left[\begin{array}{l}
\{\text{de}\}_{Prep} \text{ (from)} \\
\text{ARGSTR} = \left[\begin{array}{l} \text{P-ARG}_1 = r \\ \text{ARG}_1 = w = [\{\text{entidade concreta}\}_N \text{ (physical entity)}] \end{array}\right] \\
\text{QUALIA} = [\text{FORMAL} = \text{from}(r, \boxed{1}, w), \boxed{4}\ \neg\mathbf{at(e_3,} \boxed{1}\mathbf{, w))}]
\end{array}\right] \\
\text{D-ARG}_1 = v = \left[\begin{array}{l} \{\text{para}\}_{Prep} \text{ (to)} \\ \text{ARGSTR} = z = [\{\text{entidade concreta}\}_N \text{ (physical entity)}] \end{array}\right]
\end{array}\right] \\
\text{EVENTSTR} = & \left[\begin{array}{l} E_1 = e_1 : \text{process} \\ E_2 = \boxed{4} \\ \text{RESTR} = <\alpha \\ \text{HEAD} = \boxed{4} \end{array}\right] \\
\text{QUALIA} = & \left[\begin{array}{l} \text{AGENTIVE} = \text{tirar}(e_1, x, y, u) \\ \text{FORMAL} = \boxed{4}, \text{at}(e_3, \boxed{1}, z) \end{array}\right]
\end{array}\right]
$$

As stipulated in chapter 4, transitions are composed of two subevents: a preparatory process and a final state, temporally ordered so that the process precedes the final state (<α). The specification of GOAL and SOURCE is also related to this event structure, since the final state subevent is necessarily coincident with the state included in the semantic content of these components, as the AVMs in (23) and (24) illustrate.

(24) {deitar-se} (≅ lie down)

$$
\left[\begin{array}{ll}
\text{ARGSTR} = & \left[\begin{array}{l}
\text{P-ARG}_1 = e: [\{\text{mover} - \text{se}\}_V \text{ (move oneself)}] \\
\text{ARG}_1 = x = \boxed{1}\ [\{\text{entidade animada}\}_N \text{ (animated entity)}] \\
\text{D-ARG}_1 = u = [[\text{initial position}]] \\
\text{S-ARG}_1 = v = \left[\begin{array}{ll}
[\text{into a horizontal position}] & \\
\text{ARGSTR} = & \begin{bmatrix} \text{P-ARG}_1 = e \\ \text{ARG}_1 = \boxed{1} \end{bmatrix} \\
\text{EVENTSTR} = & \begin{bmatrix} E_1 = \boxed{3}\, e_2 : \text{state} \\ \text{HEAD} = e_2 \end{bmatrix} \\
\text{QUALIA} = & \begin{bmatrix} \text{FORMAL} = \neg\text{horizontal\_position}(e_1, \boxed{1}), \\ \boxed{4}\ \textbf{in\_horizontal\_position}(e_2, \boxed{1}) \end{bmatrix}
\end{array}\right]
\end{array}\right] \\
\text{EVENTSTR} = & \begin{bmatrix} E_1 = e_1 : \text{process} \\ E_2 = \boxed{3} \\ \text{RESTR} = <\alpha \\ \text{HEAD} = \boxed{3} \end{bmatrix} \\
\text{QUALIA} = & \begin{bmatrix} \text{AGENTIVE} = \text{deitar-se}(e_1, x, v) \\ \text{FORMAL} = \text{in}(e_3, x, u), \boxed{4} \end{bmatrix}
\end{array}\right]
$$

As discussed in chapter 3, the specification of SOURCE and GOAL components can result in accomplishment or achievement type events. According the analysis presented, both accomplisment and achievement type events are complex events composed of a process and a final state.

This distinction between these types of events is reflected in the head of the event structure: accomplishment denoting verbs have as head subevent the event argument referring to the process; achievement denoting verbs have as head subevent the event argument referring to the final state.

Note that although the punctual aspect of the movement verbs specifying GOAL or SOURCE, demonstrated by the readings of the minimal sentences in which they can occur (i.e. simple sentences realizing only the true arguments of the verb, for instance "O rapaz saiu da sala" (≅ the boy exited the room)), and by the contexts of occurrence with *for* adverbials, for instance "O rapaz saiu da sala durante 2 horas" (≅ the boy was out of the room for 2 hours) points out to a prominence of the final state in these verbs meaning, the aspectual value of the sentence in which these verbs occur can be contextually altered, forcing the process prominence:

(25) a. O homem saiu do **labirinto em 3 horas**.

(The man exited the **maze in 3 hours**) ≅ the man took 3 hours in the process of exiting the maze

b. O homem **está a sair** do labirinto.

(The man **is exiting** the maze) ≅ the man is in the process of exiting the maze

When it comes to regularities concerning syntactic realization, GOAL and SOURCE denoting arguments are typically expressed by prepositional phrases in change of location verbs and by adverbial phrases in change of position verbs:

(26) {*pôr-se*}$_V$ (≅ move oneself **into a given position**)

O homem pôs-se **de lado/de costas/de joelhos**/… (The man put himself **sideways/on his back/on his knees**/…)

6.2.2.2 DIRECTION

The semantic component DIRECTION is usually lexicalized, corresponding to the adding of a shadow argument to the argument structure of the verbs, corresponding to the DIRECTION lexicalized by the prepositional phrase introducing other components.

The nodes {avançar}$_V$ (≅ move forward), {recuar}$_V$ (≅ move backward), {subir}$_V$ (≅ move up), {descer}$_V$ (≅ move down), {aproximar}$_V$ (≅ move closer to), {afastar}$_V$ (≅ move away from), {entrar}$_V$ (≅ move in, enter), {sair}$_V$ (≅ move out), {subir 2}$_V$ (≅ go to on top of), {levantar}$_V$ (≅ place upwards), {baixar}$_V$ (≅ place down) exemplify lexicalization of DIRECTION.

On its own, the lexicalization of DIRECTION does not result in Aktionsart shifts, i.e., hyponym verbs that only lexicalize DIRECTION maintain the event structure of their hyperonym:

(27) $\{\text{avançar}\}_V$ (≅ move forward)

$$\begin{bmatrix} \text{ARGSTR} = \begin{bmatrix} \text{P-ARG1} = e: \begin{bmatrix} \{\text{mover-se}\}_V \text{ (move oneself)} \\ \ldots \\ \text{EVENTSTR} = \begin{bmatrix} \mathbf{E1 = e1 : process} \\ \mathbf{HEAD = e1} \end{bmatrix} \\ \ldots. \end{bmatrix} \\ \text{ARG1} = x = [\{\text{entidade}\}_N \text{ (entity)}] \\ \mathbf{S\text{-}ARG1 = t} = \begin{bmatrix} \mathbf{\{para\}_{Prep} \text{ (in the direction of)}} \\ \mathbf{ARGSTR} = \begin{bmatrix} \mathbf{P\text{-}ARG1 = r} \\ \mathbf{ARG1 = s = \{frente\}_N \text{ (front)}} \end{bmatrix} \\ \mathbf{QUALIA} = [\mathbf{FORMAL = to(r, \boxed{2}, s)}] \end{bmatrix} \\ \text{D-ARG1} = u = [\{\text{de}\}_{Prep} \text{ (from)} \ldots] \\ \text{D-ARG2} = v = [\{\text{para}\}_{Prep} \text{ (to)} \ldots] \end{bmatrix} \\ \text{EVENTSTR} = \begin{bmatrix} \mathbf{E1} = \boxed{2}\ \mathbf{e1 : process} \\ \mathbf{HEAD = e1} \end{bmatrix} \\ \text{QUALIA} = \begin{bmatrix} \text{AGENTIVE} = \text{avançar}(e1, x, t) \\ \text{FORMAL} = \text{at}(e3, x, z), \text{at}(e3, x, w) \end{bmatrix} \end{bmatrix}$$

(28) $\{\text{subir 2}\}_V$ (≅ go to on top of)

$$\begin{bmatrix} \text{ARGSTR} = \begin{bmatrix} \text{P-ARG1} = e: \begin{bmatrix} \{\text{ir}\}_V \text{ (go)} \\ \ldots \\ \text{EVENTSTR} = \begin{bmatrix} \mathbf{E1 = e1 : process} \\ \mathbf{E2 = e2 : state} \\ \mathbf{RESTR} = < \alpha \\ \mathbf{HEAD = e2} \end{bmatrix} \\ \ldots. \end{bmatrix} \\ \text{ARG1} = x = \boxed{1}[\{\text{entidade animada}\}_N \text{ (animated entity)}] \\ \text{D-ARG1} = u = [\{\text{de}\}_{Prep} \text{ (from)} \ldots] \\ \mathbf{ARG2 = v} = \begin{bmatrix} \mathbf{\{para\ cima\ de\}_{Prep}} \\ \mathbf{ARGSTR} = \begin{bmatrix} \mathbf{P\text{-}ARG1 = r} \\ \mathbf{ARG1 = w = [entidade\ concreta\ \text{(physical entity)}]} \end{bmatrix} \\ \mathbf{QUALIA} = [\mathbf{FORMAL = to(r, \boxed{1}, w)}] \end{bmatrix} \end{bmatrix} \\ \text{EVENTSTR} = \begin{bmatrix} \mathbf{E1 = e1 : process} \\ \mathbf{E2 = e2 : state} \\ \mathbf{RESTR} = < \alpha \\ \mathbf{HEAD = e2} \end{bmatrix} \\ \text{QUALIA} = \begin{bmatrix} \text{AGENTIVE} = \text{subir}(e1, x, v) \\ \text{FORMAL} = \text{at}(e3, x, z), \text{at}(e2, x, w) \end{bmatrix} \end{bmatrix}$$

DIRECTION denoting arguments can correspond to prepositional phrases that modify the event, in (27) above, or to more specific goal expressions, in (28).

One of the most common verbs of movement in Portuguese is the change of location verb {vir}$_V$ (≅ come; move in the direction of the speaker's reference location), that lexicalizes DIRECTION, also requiring the realization of a goal location denoting argument.

(29) Ele veio para a escola/aqui/#lá.
(He came to the school/here/#there)

Interestingly, and given that the GOAL location corresponds to the location of the speaker (*here*) the verb *vir*, as its equivalent in English, the verb *come*, occurs very frequently only with SOURCE and PATH denoting arguments since the GOAL location is established situationally:

(30) a. Ele veio [do escritório]$_{\text{SOURCE}}$ *[para aqui]*$_{GOAL}$
(He came [from the office]$_{\text{SOURCE}}$ *[to here]*$_{GOAL}$)

b. Ele veio [pelo caminho mais curto]$_{\text{PATH}}$ *[para aqui]*$_{GOAL}$
(He went [through the shortest way]$_{\text{PATH}}$ *[to here]*$_{GOAL}$)

(31) {vir}V (≅ come)

$$
\begin{bmatrix}
\text{ARGSTR} = \begin{bmatrix}
\text{P-ARG}_1 = \text{e} : \begin{bmatrix} \{\text{mover-se}\}_{\text{V}}\ (\text{move oneself}) \\ \ldots \end{bmatrix} \\
\text{ARG}_1 = \text{x} = \begin{bmatrix} \{\text{entidade animada}\}_{\text{N}}\ (\text{animated entity}) \end{bmatrix} \\
\text{D-ARG}_1 = \text{u} = \begin{bmatrix} \{\text{de}\}_{\text{Prep}}\ (\text{from}) \ldots \end{bmatrix} \\
\textbf{S-ARG}_\mathbf{1} = \mathbf{t} = \begin{bmatrix} \textbf{\{em direcção a\}}_{\textbf{Prep}}\ (\text{in the direction of}) \\ \textbf{ARGSTR} = \begin{bmatrix} \textbf{P-ARG}_\mathbf{1} = \mathbf{r} \\ \textbf{ARG1} = \mathbf{w} = \boxed{2}\,\textbf{[speaker's location]} \end{bmatrix} \\ \textbf{QUALIA} = \begin{bmatrix} \textbf{FORMAL} = \textbf{to(r,}\boxed{1}\textbf{,w)} \end{bmatrix} \end{bmatrix} \\
\textbf{ARG}_\mathbf{2} = \mathbf{v} = \begin{bmatrix} \textbf{\{para\}}_{\textbf{Prep}}\ (\text{to}) \\ \textbf{ARGSTR} = \begin{bmatrix} \textbf{P-ARG}_\mathbf{1} = \mathbf{r} \\ \textbf{ARG1} = \mathbf{z} = \boxed{2} \end{bmatrix} \\ \textbf{QUALIA} = \begin{bmatrix} \textbf{FORMAL} = \textbf{to(r,}\boxed{1}\textbf{,}\boxed{2}\textbf{), at(e}_\mathbf{1}\textbf{, x,}\boxed{2}\textbf{)} \end{bmatrix} \end{bmatrix}
\end{bmatrix} \\
\text{EVENTSTR} = \begin{bmatrix}
\mathbf{E_1} = \boxed{1}\,\mathbf{e_1 : process} \\
\mathbf{E2} = \mathbf{e_2 : state} \\
\textbf{RESTR} = <\alpha \\
\textbf{HEAD} = \mathbf{e2}
\end{bmatrix} \\
\text{QUALIA} = \begin{bmatrix}
\text{AGENTIVE} = \text{vir}(\text{e}_1,\text{x},\text{u},\text{t},\text{v}) \\
\text{FORMAL} = \text{at}(\text{e}_3,\text{x},\text{z}), \text{at}(\text{e}_2,\text{x},\text{w})
\end{bmatrix}
\end{bmatrix}
$$

Note that true arguments may be syntactically omitted if recoverable by context, which is the case here: if one comes from the office, one necessarily moves to the location of the speaker, moving in that direction; if one comes through a given path, one necessarily follows the direction of that path to the speaker's location.

6.2.2.3 PATH

The specification of PATH, although not too frequent, has several occurrences in the class of change of location verbs. As discussed in chapter 3, it results in the increase of the number of true arguments in the argument structure of the verb that denotes it, in comparison with its hyperonym.

The lexicalization of PATH requires a new true argument denoting the reference object (GROUND) with regards to which the PATH is established, realized by a NP argument (as illustrated in (35) below). The expression of the GROUND object disallows the expression of SOURCE and GOAL locations, since these are established by the PATH denoting argument with regard to the GROUND. Note that verbs that lexicalize both SOURCE and GOAL components, such as *atravessar* (≅ cross; move from one side to the other of something), also require a GROUND (or obstacle) denoting argument and also do not license the expression of SOURCE and GOAL locations:

(32) a. #O homem circum-navegou o continente [da ponta este]$_{SOURCE}$ [para a ponta este]$_{GOAL}$.
(The man circumnavigated the continent [from the east point]$_{SOURCE}$ [to the east point]$_{PATH}$)

(33) a. #O homem atravessou o continente [da ponta este]$_{SOURCE}$ [para a ponta oeste]$_{GOAL}$.
(The man crossed the continent [from the east point]$_{SOURCE}$ [to the west point]$_{PATH}$)

Consequently, the verbs that lexicalize PATH denote accomplishment type events. And so, the presence of such an argument, GROUND, is directly related to Aktionsart shifts.

(34) a. O homem circum-navegou o continente em 10 dias.
(The man circumnavigated the continent in 10 days)

b. O homem está (agora) a circum-navegar o continente → o homem (ainda) não circum-navegou o continente.
(The man is (now) circumnavigating the continent → the man has not (yet) circumnavigated the continent)

(35) {circum-navegar} (≅ circum-navigate)

$$\left[\begin{array}{l}
\text{ARGSTR} = \left[\begin{array}{l}
\text{P-ARG1} = \text{e}: \left[\{\text{navegar}\}_{\text{V}}\ (\text{navigate})\right] \\
\text{ARG1} = \text{x} = \left[\{\text{pessoa}\}_{\text{N}}\ (\text{person})\right] \\
\text{S-ARG1} = \text{t} = \left[\begin{array}{l}
\{\text{à volta de}\}_{\text{Prep}}\ (\text{around}) \\
\text{ARGSTR} = \left[\begin{array}{l} \text{P-ARG1} = \text{r} \\ \text{ARG2} = \boxed{2} \end{array}\right] \\
\text{QUALIA} = \left[\text{FORMAL} = \text{around}(\text{r}, \boxed{1}, \boxed{2})\right]
\end{array}\right] \\
\text{ARG2} = \text{y} = \boxed{2} \left[\begin{array}{l} \{\text{ilha}\}_{\text{N}}\ (\text{island}) \\ \ldots \end{array}\right] \vee \left[\begin{array}{l} \{\text{continente}\}_{\text{N}}\ (\text{continent}) \\ \ldots \end{array}\right]
\end{array}\right] \\
\text{EVENTSTR} = \left[\begin{array}{l}
\text{E1} = \boxed{1}\ \text{e1}: \text{process} \\
\text{E2} = \text{e2}: \text{state} \\
\text{RESTR} = <\alpha \\
\text{HEAD} = \text{e2}
\end{array}\right] \\
\text{QUALIA} = \left[\begin{array}{l}
\text{AGENTIVE} = \text{circum-navegar}(\text{e1}, \text{x}, \text{t}, \text{y}) \\
\text{FORMAL} = \neg\text{at_the_end_of_the_path}(\text{e3}, \text{x}, \text{t}), \\
\qquad\qquad \text{at_the_end_of_the_path}(\text{e2}, \text{x}, \text{t})
\end{array}\right]
\end{array}\right]$$

The case of incorporation of restrictions on the PATH component results only in the increase of a new true argument corresponding to a path denoting argument expressed by a prepositional phrase, typically introduced by *por* (≅ through) (in (36)).

(36) {retroceder}V (≅ move back through the same path)

$$\left[\begin{array}{l}
\text{ARGSTR} = \left[\begin{array}{l}
\text{P-ARG1} = \text{e}: \left[\{\text{recuar}\}_{\text{V}}\ (\text{move backwards})\right] \\
\text{ARG1} = \text{x} = \left[\{\text{entidade animada}\}_{\text{N}}\ (\text{animated entity})\right] \\
\text{D-ARG1} = \text{u} = \left[\{\text{de}\}_{\text{Prep}}\ (\text{from}) \ldots\right] \\
\text{D-ARG2} = \text{v} = \left[\{\text{para}\}_{\text{Prep}}\ (\text{to}) \ldots\right] \\
\text{ARG2} = \text{t} = \left[\begin{array}{l}
\{\text{por}\}_{\text{Prep}}\ (\text{through}) \\
\text{ARGSTR} = \left[\begin{array}{l} \text{P-ARG1} = \text{r} \\ \text{ARG1} = \text{s} = \left[\begin{array}{l} \{\text{via}\}_{\text{N}}\ (\text{path}) \\ \ldots \end{array}\right] \vee \left[\begin{array}{l} \{\text{local}\}_{\text{N}}\ (\text{location}) \\ \ldots \end{array}\right] \end{array}\right] \\
\text{QUALIA} = \left[\text{FORMAL} = \text{through}(\text{r}, \boxed{1}, \text{s})\right]
\end{array}\right]
\end{array}\right] \\
\text{EVENTSTR} = \left[\begin{array}{l}
\text{E1} = \boxed{1}\ \text{e1}: \text{process} \\
\text{HEAD} = \text{e1}
\end{array}\right] \\
\text{QUALIA} = \left[\begin{array}{l}
\text{AGENTIVE} = \text{retroceder}(\text{e1}, \text{x}, \text{t}) \\
\text{FORMAL} = \text{at}(\text{e2}, \text{x}, \text{z}), \text{at}(\text{e3}, \text{x}, \text{w})
\end{array}\right]
\end{array}\right]$$

In these cases, the verbs can co-occur with SOURCE and GOAL denoting default arguments:

(37) a. O homem retrocedeu pela pista [da meta]$_{SOURCE}$ [para o primeiro obstáculo]$_{GOAL}$.
(The man moved back through the track [from the finish line]$_{SOURCE}$ [to the first obstacle]$_{PATH}$)

Also, the expression of both SOURCE and GOAL default arguments may license the omission of the PATH denoting argument, since it provides the contextual information that allows its recovery:

(38) a. O homem retrocedeu [da meta]$_{SOURCE}$ [para o primeiro obstáculo]$_{GOAL}$.
(The man moved back [from the finish line]$_{SOURCE}$ [to the first obstacle]$_{PATH}$)

From the data explained above, the lexicalization of PATH establishes two distinct patterns:

i) Verbs that lexicalize PATH require the realization of a true argument express by a NP, do not inherit GOAL and SOURCE default arguments and denote accomplishment type events. Examples:

(39) a. {circundar}$_V$ [≅ circumbulate; move around something]

b. {contornar}$_V$ [≅circumvent; move near the limits or boundaries of something]

c. {circum-navegar}$_V$ [≅circumnavigate; navigate around a mass of land]

ii) verbs that incorporate restrictions on the PATH component (in the cases observed some simultaneously lexicalizing MANNER) require a PATH denoting argument and denote activity type events:

(40) a. {circular, transitar, andar}$_V$ [≅circulate; move freely and usually through a given space/way].

b. {circular}$_V$ [≅ circulate; move cyclically, through a given path]

c. {seguir}$_V$ [≅move along; move for a while through a given path]

As mentioned in chapter 3, these two patterns may also be related to the lexicalization of MANNER, since the MANNER specified in the content of the verbs presented above concerns the duration of the event, whether by establishing its iterativity (*usually*, in (40)a and *cycically* in (40)b) whether by establishing that the event denoted extends in time (*for a while*, in (40)c).

6.2.2.4 MANNER

MANNER lexicalization is transversal across semantic domains. That is, contrary to SOURCE, GOAL, DIRECTION and PATH, it does not occur only in the class of verbs of movement. Typically, the specification of MANNER corresponds to the increase of a shadow argument in the argument structure of the verb.

The MANNER component may correspond to an adverbial expression (in (41)) or to the combination of more than one expression (in (42)):

(41) $\{\text{cambalear}\}_V$ ($\cong$ walk unsteadly)

$$\begin{bmatrix} \text{ARGSTR} = \begin{bmatrix} \text{P-ARG}_1 = \text{e} : [\{\text{andar}\}_V \text{ (walk)}] \\ \text{ARG}_1 = \text{x} = [\{\text{animal}\}_N \text{ (animal)}] \\ \\ \text{S-ARG}_1 = \text{t} = \begin{bmatrix} \mathbf{\{tropegamente\}_{Adv}} \text{ (unsteadly)} \\ \text{ARGSTR} = \begin{bmatrix} \text{P-ARG}_1 = \text{m} \\ \text{ARG}_1 = \boxed{1} \end{bmatrix} \\ \text{QUALIA} = [\text{FORMAL} = \text{unsteadly(m, } \boxed{1}\text{)}] \end{bmatrix} \\ \text{D-ARG}_1 = \text{u} = [\{\text{de}\}_{Prep} \text{ (from)} \ldots] \\ \text{D-ARG}_2 = \text{v} = [\{\text{para}\}_{Prep} \text{ (to)} \ldots] \end{bmatrix} \\ \text{EVENTSTR} = \begin{bmatrix} \text{E}_1 = \boxed{1}\, \text{e}_2 : \text{process} \\ \text{HEAD} = \boxed{1} \end{bmatrix} \\ \text{QUALIA} = \begin{bmatrix} \text{AGENTIVE} = \text{cambalear(e}_1\text{,x,t)} \\ \text{FORMAL} = \text{at(e}_2\text{, x,z), at(e}_3\text{, x,w)} \end{bmatrix} \end{bmatrix}$$

(42) $\{\text{estremecer}\}_V$ ($\cong$ tremble suddenly and ocasionally)

$$\begin{bmatrix} \text{ARGSTR} = \begin{bmatrix} \text{P-ARG}_1 = \text{e} : [\{\text{tremer}\}_V \text{ (tremble)}] \\ \text{ARG}_1 = \text{x} = [\{\text{animal}\}_N \text{ (animal)}] \\ \ldots \\ \text{S-ARG}_1 = \text{t} = \begin{bmatrix} \mathbf{\{subitamente\}_{Adv}} \\ \text{(suddenly)} \\ \text{ARGSTR} = \begin{bmatrix} \text{P-ARG}_1 = \text{m} \\ \text{ARG}_1 = \boxed{1} \end{bmatrix} \\ \text{QUALIA} = \begin{bmatrix} \text{FORMAL} = \\ \text{suddenly(m, } \boxed{1}\text{)} \end{bmatrix} \end{bmatrix} \wedge \begin{bmatrix} \mathbf{\{ocasionalmente\}_{Adv}} \\ \text{(occasionally)} \\ \text{ARGSTR} = \begin{bmatrix} \text{P-ARG}_1 = \text{m} \\ \text{ARG}_1 = \boxed{1} \end{bmatrix} \\ \text{QUALIA} = \begin{bmatrix} \text{FORMAL} = \\ \text{occasionally(m, } \boxed{1}\text{)} \end{bmatrix} \end{bmatrix} \end{bmatrix} \\ \text{EVENTSTR} = \begin{bmatrix} \text{E}_1 = \boxed{1}\, \text{e}_2 : \text{process} \\ \text{HEAD} = \boxed{1} \end{bmatrix} \\ \text{QUALIA} = [\text{AGENTIVE} = \text{estremecer(e}_1\text{,x,t)}] \end{bmatrix}$$

Note, though, that this is not always the case. The MANNER notion may not correspond to any lexicalized concept. In these cases, we have chosen to represent it in the AVMs using the paraphrase that expresses it, although, *a priori*, this semantic content will not be expressed in the wordnet, as it will be discussed in the next chapter.

(43) {abanar}$_V$ (≅ shake)

$$\begin{bmatrix} \text{ARGSTR} = \begin{bmatrix} \text{P-ARG1} = e : \begin{bmatrix} \{\text{mexer}\}_V(\text{change position}) \\ \ldots \end{bmatrix} \\ \text{ARG1} = x = [\{\text{entidade}\}_N\ (\text{entity})] \\ \text{ARG2} = y = [\{\text{objecto}\}_N\ (\text{object})] \\ \ldots \\ \text{S-ARG1} = w = [\textbf{[with impulses in different directions]}] \end{bmatrix} \\ \text{EVENTSTR} = \begin{bmatrix} \text{E1} = e1 : \text{process} \\ \text{HEAD} = e1 \end{bmatrix} \\ \text{QUALIA} = \begin{bmatrix} \text{AGENTIVE} = \text{abanar}(e1, x, y, w) \\ \text{FORMAL} = \text{in}(e2, y, u), \text{in}(e3, y, v) \end{bmatrix} \end{bmatrix}$$

Although not occurring frequently in the set of verbs analyzed, MANNER can also correspond to instrument denoting arguments, with adverbial sense since they modify the event, syntactically expressed by a prepositional phrase headed by the preposition *com* (≅ with) or *de* (≅ by), in (44) and (45) respectively.

(44) {içar}$_V$ (≅ move up with ropes or cable)

$$\begin{bmatrix} \text{ARGSTR} = \begin{bmatrix} \text{P-ARG1} = e : [\{\text{erguer}\}_V\ (\text{move up})] \\ \text{ARG1} = x = [\{\text{pessoa}\}_N\ (\text{person})] \\ \text{ARG2} = y = [\{\text{objecto}\}_N\ (\text{object})] \\ \text{S-ARG1} = t = \begin{bmatrix} \textbf{\{com\}}_{\textbf{Prep}}\ \textbf{(with)} \\ \text{ARGSTR} = \begin{bmatrix} \text{P-ARG1} = r \\ \text{ARG2} = s = \begin{bmatrix} \{\text{corda}\}_N\ (\text{rope}) \\ \ldots \end{bmatrix} \vee \begin{bmatrix} \{\text{cabo}\}_N\ (\text{cable}) \\ \ldots \end{bmatrix} \end{bmatrix} \\ \text{QUALIA} = [\text{FORMAL} = \text{with}(r, \boxed{1}, s)] \end{bmatrix} \\ \text{D-ARG1} = u = [\{\text{de}\}_{\text{Prep}}\ (\text{from}) \ldots] \\ \text{D-ARG2} = v = [\{\text{para}\}_{\text{Prep}}\ (\text{to}) \ldots] \end{bmatrix} \\ \text{EVENTSTR} = \begin{bmatrix} \text{E1} = \boxed{1}\, e1 : \text{process} \\ \text{HEAD} = e1 \end{bmatrix} \\ \text{QUALIA} = \begin{bmatrix} \text{AGENTIVE} = \text{içar}(e1, x, y, t) \\ \text{FORMAL} = \text{at}(e2, x, z), \text{at}(e3, y, w) \end{bmatrix} \end{bmatrix}$$

(45) {acarretar}$_V$ (≅ transport by car)

$$\left[\begin{array}{l}
\text{ARGSTR} = \left[\begin{array}{l}
\text{P-ARG}_1 = e: \left[\{\text{transportar}\}_V \text{ (transport) }\right] \\
\text{ARG}_1 = x = \left[\{\text{pessoa}\}_N \text{ (person)}\right] \\
\text{ARG2} = y = \left[\{\text{objecto}\}_N \text{ (object)}\right] \\
\text{S-ARG}_1 = t = \left[\begin{array}{l}
\mathbf{\{de\}_{Prep}\ (by)} \\
\text{ARGSTR} = \left[\begin{array}{l}
\text{P-ARG}_1 = r \\
\text{ARG2} = s = \left[\begin{array}{l} \{\text{carro}\}_N \text{ (car)} \\ \ldots \end{array}\right]
\end{array}\right] \\
\text{QUALIA} = \left[\text{FORMAL} = \text{with}(r, \boxed{1}, s)\right]
\end{array}\right] \\
\text{D-ARG}_1 = u = \left[\{\text{de}\}_{Prep} \text{ (from)} \ldots\right] \\
\text{D-ARG}_2 = v = \left[\{\text{para}\}_{Prep} \text{ (to)} \ldots \right]
\end{array}\right] \\
\text{EVENTSTR} = \left[\begin{array}{l}
\text{E}_1 = \boxed{1}\, e_1 : \text{process} \\
\text{HEAD} = e_1
\end{array}\right] \\
\text{QUALIA} = \left[\begin{array}{l}
\text{AGENTIVE} = \text{acarretar}(e_1, x, y, t) \\
\text{FORMAL} = \text{at}(e_2, x, z),\ \text{at}(e_3, y, w)
\end{array}\right]
\end{array}\right]$$

Finally, MANNER components can also correspond to a verbal expression. For instance, the verb *esbracejar* denotes the event of moving oneself (changing oneself position) by agitating one's arms vigorously. In this way, the event denoted by {esbracejar}$_V$ entails, at a deeper level (see discussion of process type event properties in chapter 4, section 4.1.1.2) the subevent of agitating one's arms, that also determines MANNER.

(46) {esbracejar}$_V$ (≅ moving oneself by agitating one's arms vigorously)

$$\left[\begin{array}{l}
\text{ARGSTR} = \left[\begin{array}{l}
\text{P-ARG}_1 = e: \left[\{\text{mover-se}\}_V \text{ (move oneself) }\right] \\
\text{ARG}_1 = x = \boxed{1} \left[\begin{array}{l}
\{\text{pessoa}\}_N \text{ (person)} \\
\text{ARGSTR} = \left[\text{P-ARG1} = x : [\{\text{animal}\}_N]\right] \\
\text{QUALIA} = \left[\text{CONSTITUTIVE} = \text{has_part}(x, y : \boxed{2}\, \text{arm})\right]
\end{array}\right] \\
\text{D-ARG}_1 = u = \left[[\text{initial position}]\right] \\
\text{D-ARG}_2 = v = \left[[\text{final position}]\right] \\
\text{S-ARG}_1 = t = \left[\begin{array}{l}
\mathbf{\{agitar\}_v\ (agitate)} \\
\text{ARGSTR} = \left[\begin{array}{l}
\text{P-ARG}_1 = e \\
\text{ARG}_1 = \boxed{1} \\
\text{ARG}_2 = \boxed{2}
\end{array}\right] \\
\text{EVENTSTR} = \left[\begin{array}{l}
\text{E1} = \boxed{3}\, e_1 : \text{process} \\
\text{HEAD} = e_1
\end{array}\right] \\
\text{QUALIA} = \left[\begin{array}{l}
\text{FORMAL} = \\
\quad \text{agitar}(e_1, \boxed{1}, \boxed{2})
\end{array}\right]
\end{array}\right] \wedge \left[\begin{array}{l}
\mathbf{\{vigorosamente\}_{Adv}} \\
\mathbf{(vigorously)} \\
\text{ARGSTR} = \left[\begin{array}{l}
\text{P-ARG}_1 = m \\
\text{ARG}_1 = \boxed{3}
\end{array}\right] \\
\text{QUALIA} = \left[\begin{array}{l}
\text{FORMAL} = \\
\text{vigorously}(m, \boxed{3})
\end{array}\right]
\end{array}\right]
\end{array}\right] \\
\text{EVENTSTR} = \left[\begin{array}{l}
\text{E}_1 = e_1 : \text{process} \\
\text{HEAD} = e_1
\end{array}\right] \\
\text{QUALIA} = \left[\begin{array}{l}
\text{AGENTIVE} = \text{esbracejar}(e_1, x, t) \\
\text{FORMAL} = \text{in}(e_2, x, u),\ \text{in}(e_3, x, v)
\end{array}\right]
\end{array}\right]$$

Lexicalization of MANNER is not necessarily associated to Aktionsart shifts, but it can reinforce the prominence of process subevents in transition denoting verbs. The following verbs seem to exemplify this phenomenon: $\{drenar\}_V$ (≅ drain; take liquids from organic cavities using a drain); $\{despejar, verter\}_V$ (≅ spill; take the content of a container by inverting its position, through gravity action); $\{escoar\ 1\}_V$ (≅ take liquids, using channels and openings); $\{extrair\}_V$ (≅ extract; take with effort, something fastened), $\{bombear\}_V$ (≅ pump; take liquids using a pump). These verbs are hyponyms of the achievement denoting verb $\{retirar\}_V$ (≅ take out) and incorporate restrictions on the semantic component SOURCE.

The lexicalization of MANNER by hyponym verbs seems to result in an Aktionsart shift, as the sentences below show:

(47) a. O homem retirou a caixa da sala durante 2 minutos.
(The man took the box from the room for 2 minutes) (≅ the box was out of the room during 2 minutes)

(48) a. O homem drenou o abcesso durante 2 minutos.
(The man drained the abcess for 2 minutes) (≅ the man was 2 minutes in the process of draining, and not the liquid was out of the abcess during 2 minutes)

b. O homem despejou/verteu a água do copo durante 2 minutos.
(The man poured the water from the glass for 2 minutes) (≅ the man was 2 minutes in the process of pouring, and not the water was out of the glass during 2 minutes)

c. O homem escoou a água do tanque durante 2 minutos.
(The man drained the water from the tank for 2 minutes) (≅ the man was 2 minutes in the process of draining, and not the water was out of the tank during 2 minutes)

d. O homem extraiu a bala do corpo durante 2 minutos.
(The man extracted the bullet from the body for 2 minutes) (≅ the man was 2 minutes in the process of extracting, and not the bullet was out of the body during 2 minutes)

e. O homem bombeou a água do tanque durante 2 minutos.
(The man pumped the water from the tank for 2 minutes) (≅ the man was 2 minutes in the process of pumping, and not the water was out of the tank during 2 minutes)

However, the sentence in (49) shows that the massive aspect of the FIGURE in these cases also contributes to the accomplishment reading:

(49) a. O homem retirou a água da sala durante 2 minutos.
(The man took the water from the room for 2 minutes) (≅ the man was 2 minutes in the process of taking the water, and not the water was out of the room during 2 minutes)

For this reason, it is not possibe to undoubtly relate lexicalization of MANNER in verbs of movement and Aktionsart shifts from achievements to accomplishments.

6.2.2.5 FIGURE

As briefly described in chapter 3, FIGURE is rarely, or not all, lexicalized. Portuguese verbs of movement incorporate restrictions on the FIGURE component but do not lexicalize it. Also, the cases where the hyponym distinguishes itself from the hyperonym meaning by the specification of FIGURE alone are not common. For these reasons, there seems to be no salient semantic or syntactic pattern of behavior associated to this element.

FIGURE is reflected in selection constraints on the type of arguments (second true argument in the case of two-place verbs; first true argument in the case of one-place verbs). However, in cases such as the one presented in (48) and (49) above, the semantic properties of FIGURE referring arguments can be related to other specificities of the hyponym verb.

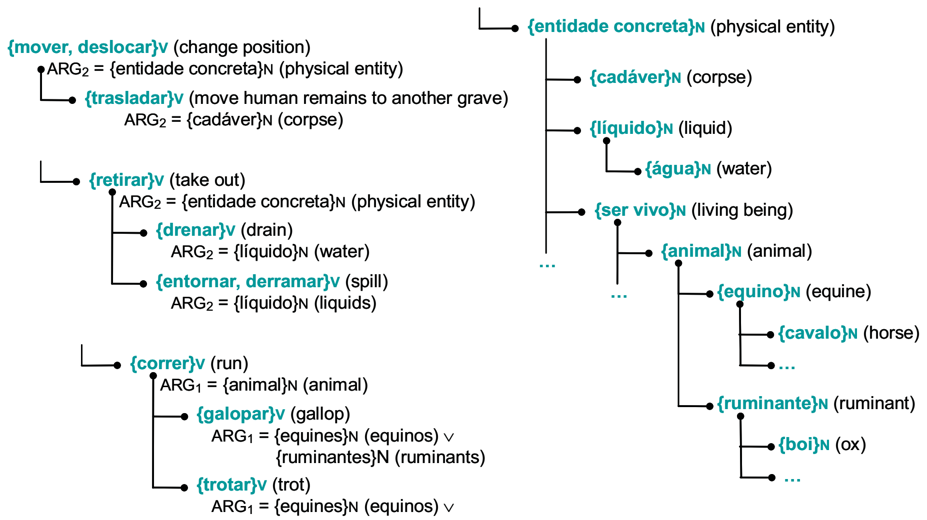
Consider, for instance, the verb $\{dispor\}_V$ (≅ arrange; put objects of a set according to a given logic), hyponym of the verb $\{pôr, colocar\}_V$ (≅ put). The specification of the type of entities that can be argument of this verb as a set of entities (whether expressed lexically by a item that denotes a set such as *class* or *group*; whether expressed morphologically by the plural: *books*, *plates*, etc.) entails iterativity, in the sense that the event described by the verb $\{dispor\}_V$ (≅ arrange) refers to a complex event that includes the individual change of location of each element of the set, although, once again, this is not directly reflected in the semantic and syntactic properties of the verb.

(50) $\{\text{dispor}\}_V$ (≅ put objects of a set according to a given logic)

$$\left[\begin{array}{l}
\text{ARGSTR} = \left[\begin{array}{l}
\text{P-ARG1} = \text{e}: [\{\text{pôr, colocar}\}_{\text{V}}\ (\text{move})]\\
\text{ARG1} = \text{x} = [\{\text{entidade}\}_{\text{N}}\ (\text{entity})]\\
\textbf{ARG2} = \textbf{y} = \boxed{1}\left[\begin{array}{l} \textbf{...}\\ \textbf{QUALIA} = [\textbf{CONSTITUTIVE} = \textbf{has_element(x, y)}]\end{array}\right]\\
\text{D-ARG1} = \text{u} = \left[\begin{array}{l} \{\text{de}\}_{\text{Prep}}\ (\text{from})\\ \text{ARGSTR} = \text{z} = [\{\text{entidade concreta}\}_{\text{N}}\ (\text{physical entity})]\end{array}\right]\\
\text{S-ARG1} = \text{s} = \left[\begin{array}{l} \{\text{por}\}_{\text{Prep}}\ (\text{in [a given order]})\\ \text{ARGSTR} = \left[\begin{array}{l}\text{P-ARG1} = \text{r}\\ \text{ARG1} = \text{y} = [\{\text{ordem}\}_{\text{N}}\ (\text{order})]\end{array}\right]\\ \text{QUALIA} = [\text{FORMAL} = \text{in}(\text{r}, \boxed{5}, \text{y})]\end{array}\right]\\
\text{ARG3} = \text{w} = \left[\begin{array}{l} \{\text{em}\}_{\text{Prep}}\ (\text{in/at})\\ \text{ARGSTR} = \left[\begin{array}{l}\text{P-ARG1} = \text{e}\\ \text{ARG1} = \text{x} = [\text{entidade concreta (physical entity)}]\end{array}\right]\\ \text{EVENTSTR} = \left[\begin{array}{l}\text{E1} = \boxed{3}\,\text{e1}: \text{state}\\ \text{HEAD} = \text{e1}\end{array}\right]\\ \text{QUALIA} = [\text{FORMAL} = \boxed{4}\,\text{at}(\text{e1}, \boxed{1}, \text{x})]\end{array}\right]
\end{array}\right]\\
\text{EVENTSTR} = \left[\begin{array}{l}
\text{E1} = \boxed{5}\,\text{e1}: \text{process}\\
\text{E2} = \boxed{3}\\
\text{RESTR} = <\alpha\\
\text{HEAD} = \text{e1}
\end{array}\right]\\
\text{QUALIA} = \left[\begin{array}{l}
\text{AGENTIVE} = \text{dispor}(\text{e1}, \text{x}, \text{y}, \text{s}, \text{w})\\
\text{FORMAL} = \text{at}(\text{e3}, \boxed{1}, \text{z}), \boxed{4}
\end{array}\right]
\end{array}\right]$$

Incorporation of semantic restrictions on FIGURE may correspond to the selection of hyponyms of the argument selected by the hyperonym node. The examples in (51) below illustrate this case.

(51) Fragments of a relational net and argument correspondence



As discussed in chapter 2, section 2.2.3, the verbs *mover-se* (≅ change oneself location) and *mexer-se* (≅ change oneself position) incorporate the second argument of their hyperonym, denoting a reflex event and resulting in the replacement of the second true argument by a shadow argument.

(52) {mexer-se}$_V$ (≅ change oneself position; move oneself)

$$\begin{bmatrix} \text{ARGSTR} = \begin{bmatrix} \text{P-ARG1} = \text{e}: [\{\text{mexer}\}\text{V (move)}] \\ \text{ARG1} = \text{x} = \boxed{1}[\{\text{entidade}\}\text{N (entity)}] \\ \text{S-ARG1} = \text{y} = \boxed{1} \\ \ldots \end{bmatrix} \\ \ldots \end{bmatrix}$$

The reflexive clitic *se* associated to these verbs can be seen as a remnant of this incorporation, indicating the co-reference between the first and second arguments inherited. The syntactic realization of these shadow arguments, as reflected in the paraphrase that describes the denoted concepts, is restricted to the pronominal expression *a si próprio* (oneself). Note, however, that not all the clitics behave as reflexive clitics, allowing this realization. For instance, the verb *espreguiçar-se* (stretch out), in which the clitic is not interpreted as co-referring the first and second true arguments of an two-place hyperonym verb, since there is no such verb. For this reason, and as already established, the shadow argument is only present in the lexical entry of the direct hyponyms that first reflect this incorporation. Second level hyponyms not do reflect this semantic element.

6.2.2.6 GROUND

GROUND lexicalization, although transversal to verbs from other semantic domains, as mentioned in the beginning of this section, is not very productive in Portuguese verbs of movement. Even though, there are two salient patterns that emerge in the observed verbs: GROUND lexicalization can consist of a shadow argument directly related to FIGURE or PATH properties (see section 6.2.2.3) or it can also result in a shadow argument expressing reference location or environment, frequently expressed syntactically by specification processes.

Verbs such as {rodar, girar}$_V$ (≅ rotate, spin), {oscilar}$_V$ (≅ oscilate), {encaixar}$_V$ (≅ place in a pre-determined position, defined with respect to another object) and {revolver, remexer}$_V$ (≅rummage move changing the relative position of objects inside a given container) illustrate the first pattern. The GROUND concept in these verbs is conveyed by a shadow argument that can be syntactically realized by the prepositional expression *em relação a/relativamente a* (≅ with respect to), as the AVM in (53) demonstrates:

(53) $\{\text{oscilar}\}_V$ (≅ oscilate)

$$\left[\begin{array}{ll}
\text{ARGSTR} = & \left[\begin{array}{l}
\text{P - ARG1} = \text{e}: [\{\text{mexer}\}\text{V}(\text{move})] \\
\text{ARG1} = \text{x} = [\{\text{entidade}\}\text{V (entity)})] \\
\text{ARG1} = \text{y} = \left[\begin{array}{l} \\ \text{ARGSTR} = [\text{P - ARG1} = \text{x}: \{\text{objecto}\}\text{N (object)}] \\ \text{QUALIA} = [\text{CONSTITUTIVE} = \text{has_part}(\text{x}, \boxed{1}\,\text{y}: \text{axis})] \end{array}\right] \\
\text{S - ARG1} = \text{u} = \left[\begin{array}{l} [\text{de uma posição}]\ (\text{from one position} \\ \ldots \end{array}\right] \\
\text{S - ARG2} = \text{v} = \left[\begin{array}{l} [\text{para a posição oposta}]\ (\text{to the opposite position}) \\ \ldots \end{array}\right] \\
\text{S - ARG3} = \text{t} = \left[\begin{array}{l} \{\text{alternadamente}\}\text{Adv (alternating)}] \\ \text{ARGSTR} = \left[\begin{array}{l} \text{P - ARG1} = \text{m} \\ \text{ARG1} = \boxed{2} \end{array}\right] \\ \text{QUALIA} = [\text{FORMAL} = \text{alternating}(\text{m}, \boxed{2})\ ] \end{array}\right] \\
\text{S - ARG4} = \text{w} = \left[\begin{array}{l} \{\textbf{em relação a/relativamente a}\}\text{Prep (with respect to)} \\ \text{ARGSTR} = \left[\begin{array}{l} \text{P - ARG1} = \text{r} \\ \text{ARG2} = \boxed{2} \end{array}\right] \\ \text{QUALIA} = [\text{FORMAL} = \text{with respect to}(\text{e1}, \boxed{2}, \boxed{1})] \end{array}\right] \\
\ldots
\end{array}\right] \\
\text{EVENTSTR} = & \left[\begin{array}{l} \text{E1} = \boxed{2}\,\text{e1}: \text{process} \\ \text{HEAD} = \text{e1} \end{array}\right] \\
\text{QUALIA} = & \left[\begin{array}{l} \text{AGENTIVE} = \text{oscilar}(\text{e1}, \text{x}, \text{y}, \text{u}, \text{v}, \text{t}, \text{w}) \\ \text{FORMAL} = \text{in}(\text{e2}, \text{y}, \text{w}), \text{in}(\text{e3}, \text{y}, \text{z}) \end{array}\right]
\end{array}\right]$$

In order to explain the interdependence of the FIGURE and the object with respect to which the event is viewed (the GROUND), expressed in the lexical entry of this verb, let us consider first the meaning of the verb *oscilar*. The concept denoted by $\{\text{oscilar}\}_V$ (≅ oscilate) can be paraphrased as follows: 'move from one position to the other, in alternation, with respect to an axis'. However, the determination of the 'axis', the GROUND, is directly related to the second true argument selected by the verb, i.e., the FIGURE. For this reason, besides being an object, the second argument of the verb *oscilar* has to have as part an axis, or a part that is perceived as an axis. This specification is achieved through an underspecified representation, considering only information about the hyperonym of the argument, i.e., it has to be an hyponym of $\{\text{objecto}\}_N$ (≅ object), and about the Constitutive role of this object, has to have as part an axis.

The second pattern can be observed in verbs such as $\{\text{voar}\}_V$ (≅ fly), $\{\text{navegar}\}_V$ (≅ navigate) or $\{\text{nadar}\}_V$ (≅ swim). In these cases, the GROUND corresponds to a reference location or environment and is represented as a shadow argument, syntactically realized as a prepositional phrase introduced by the preposition *em* (≅ in), indicator of location:

(54) {voar}$_V$ (≅ fly)

$$\left[\begin{array}{ll}
\text{ARGSTR} = & \left[\begin{array}{l}
\text{P-ARG}_1 = e : [\{\text{mover-se}\}_V \text{ (move oneself)}] \\
\\
\text{ARG}_1 = x = [\{\text{entidade animada}\}_N \text{ (animated entity)}] \\
\text{S-ARG}_1 = t = \left[\begin{array}{l}
\text{[with wings]} \\
\text{ARGSTR} = \begin{bmatrix} \text{P-ARG}_1 = m \\ \text{ARG}_1 = \boxed{1} \end{bmatrix} \\
\text{QUALIA} = [\text{FORMAL} = \text{with_wings}(m, \boxed{1})]
\end{array}\right] \\
\text{S-ARG}_2 = w = \left[\begin{array}{l}
\text{em (in/at)} \\
\text{ARGSTR} = \begin{bmatrix} \text{P-ARG}_1 = e \\ \text{ARG}_1 = y = \{\text{ar}\}_N \text{ (air)} \end{bmatrix} \\
\text{EVENTSTR} = \begin{bmatrix} E_1 = e_1 : \text{state} \\ \text{HEAD} = e_1 \end{bmatrix} \\
\text{QUALIA} = [\text{FORMAL} = \text{in/at}(e_1, \boxed{1}, y)]
\end{array}\right] \\
\text{D-ARG1} = u = [\{\text{de}\}\text{Prep (from)} \ldots] \\
\text{D-ARG2} = v = [\{\text{para}\}\text{Prep (to)} \ldots]
\end{array}\right] \\
\text{EVENTSTR} = & \begin{bmatrix} E_1 = \boxed{1}\, e_1 : \text{process} \\ \text{HEAD} = e_1 \end{bmatrix} \\
\\
\text{QUALIA} = & \begin{bmatrix} \text{AGENTIVE} = \text{voar}(e_1, x, t, w) \\ \text{FORMAL} = \text{at}(e_2, x, w), \text{at}(e_3, x, z) \end{bmatrix}
\end{array}\right]$$

As expected, the realization of the shadow argument must result from specification or subtyping processes. In the example above, since the concept denoted by *voar* (fly) lexicalizes as reference location or environment the concept denoted by the node {ar}$_N$ (air), the shadow argument can be syntactically realized by any expression that denotes a specification of this concept.

(55) a. O pássaro voou no céu sereno.
(The bird flew in the quiet sky)

6.2.2.7 INTENTION

The lexicalization of INTENTION in the set of Portuguese verbs of movement is not very frequent. It usually corresponds to a shadow argument, expressed by a prepositional phrase introduced by the preposition *para/com a intenção de* (≅ to/for, with the intention of), that selects an event type complement. The verbs that illustrate this case are {embalar} (≅ rock animated entities **to get them to sleep**), {vasculhar, escarafunchar} (≅ rummage **to find something**), {espreguiçar-se} (≅ move oneself by streching the muscles, **to relax**; strecht out), {acenar} (≅ move oneself agitating the arms or hands, **to draw attention**; wave, beackon), {fugir} (≅ move away, typically fast, **to escape**; run away), and {perseguir} (≅ move after something **to catch it**; chase).

Given the particular meaning aspect it specifies, of purpose or intended goal of the event, INTENTION is also reflected in the qualia structure of these verbs, as value of the TELIC role. The AVM in (55) illustrates such a case.

(56) {embalar} (≅ rock)

$$
\left[\begin{array}{l}
\text{ARGSTR} = \left[\begin{array}{l}
\text{P - ARG}_1 = e : [\{\text{balançar}\}_V \text{ (sway)}] \\
\text{ARG}_1 = x = \boxed{1}\,[\{\text{entidade animada}\}_N \text{ (animated entity)}] \\
\text{ARG}_1 = y = \boxed{2}\,[\{\text{animal}\}_N \text{ (animal)}] \\
\text{S - ARG}_1 = w = \left[\begin{array}{l}
\{\text{para/com a intenção de}\}_{Prep} \text{ (to/with the intention of)} \\
\text{ARGSTR} = \left[\begin{array}{l}
\text{P - ARG}_1 = r \\
\text{ARG}_1 = y = \left[\begin{array}{l}
\{\text{adormecer}\}_V \text{ (put to sleep)} \\
\text{ARGSTR} = \left[\begin{array}{l} \text{P - ARG1} = e \\ \text{ARG1} = \boxed{1} \\ \text{ARG2} = \boxed{2} \end{array}\right] \\
\text{EVENTSTR} = \left[\begin{array}{l} E_1 = e_1 : \text{process} \\ E_2 = e_2 : \text{state} \\ \text{RESTR} = < \alpha \\ \text{HEAD} = e_2 \end{array}\right] \\
\text{QUALIA} = \left[\begin{array}{l} \text{FORMAL} = \neg\text{asleep}(e_3, \boxed{2}), \\ \qquad\qquad \text{asleep}(e_2, \boxed{2}) \\ \text{AG.} = \boxed{3}\,\text{adormecer}(e_2, \boxed{1}, \boxed{2}) \end{array}\right]
\end{array}\right]
\end{array}\right] \\
\text{QUALIA} = [\text{FORMAL} = \text{to}(r, \boxed{4}, y)]
\end{array}\right]
\end{array}\right] \\
\text{EVENTSTR} = \left[\begin{array}{l} E_1 = e_1 : \text{process} \\ \text{HEAD} = e_1 \end{array}\right] \\
\text{QUALIA} = \left[\begin{array}{l} \text{AGENTIVE} = \boxed{4}\,\text{embalar}(e_1, x, y, w) \\ \text{TELIC} = \boxed{3} \end{array}\right]
\end{array}\right]
$$

Note that, in spite of being a value of the Telic role, the semantic predicate expressing INTENTION, corresponding to part of the semantic content of the shadow argument, (the numerical index 3, in the AVM above) is not a subevent of the process type event denoted by the verb $\{\text{embalar}\}_V$ (≅ rock). That is, *embalar* (≅ rock) lexicalizes the concept of *with the intention of putting to sleep*, but it does not entails that *put to sleep* takes place whenever someone *rocks*. This is consistent with other values of the Telic role, in the sense that not always the purpose or intended goal of the event/object/etc is lexicalized in the lexical item whose concept entails it (see also the analysis of the verb *preparar* (≅ prepare) in Marrafa & Moura (2005), referred in chapter 4, section 4.1.1.2).

6.2.2.8 CAUSE

The last semantic component to consider in the representation of the hyponym verbs of the class of Portuguese verbs of movement is CAUSE. CAUSE lexicalization is also the most uncommon lexicalization phenomenon, occurring only in one node: the verb $\{vacilar\}_V$ (≅ tremble due to structural instability).

The lexicalization of CAUSE corresponds necessarily to a new shadow argument in the argument structure of the verb, but is also reflected in the qualia structure, namely as value of the Agentive role, since it also denotes the causal chain or origin of the event.

(57) $\{vacilar\}_V$ (≅ tremble due to/because of structural instability)

$$\begin{bmatrix} \text{ARGSTR} = \begin{bmatrix} \text{P-ARG}_1 = e: [\{\text{tremer}\}_V \text{ (tremble)}] \\ \text{ARG}_1 = x = [\{\text{objecto}\}_N \text{ (object)}] \\ \text{S-ARG}_1 = w = \begin{bmatrix} \{\text{por causa de/devido a/de}\}_{Prep} \text{ (because of)} \\ \text{ARGSTR} = \begin{bmatrix} \text{P-ARG}_1 = r \\ \text{ARG}_1 = y = [\{\text{instabilidade}\}_N \text{ (instability)}] \end{bmatrix} \\ \text{QUALIA} = [\text{FORMAL} = \boxed{2}\,\text{because\_of}(r, \boxed{1}, y)] \end{bmatrix} \end{bmatrix} \\ \text{EVENTSTR} = \begin{bmatrix} \text{E}_1 = \boxed{1}\, e_1: \text{process} \\ \text{HEAD} = e_1 \end{bmatrix} \\ \text{QUALIA} = \begin{bmatrix} \text{FORMAL} = \text{vacilar}(e_1, x, w) \\ \text{AGENTIVE} = \boxed{2} \end{bmatrix} \end{bmatrix}$$

The semantic predicate present at the qualia structure level establishes a new semantic relation, directly and necessarily related to the semantic content of the shadow argument. The prepositional phrase headed by the expression *por causa de/devido a* (≅ because of/due to), indicator of cause, expresses the relation between the process denoted by the verb and the concept denoted by the noun *instabilidade* (instability) expressing, simultaneously, the cause of the process denoted by *vacilar* (numerical index 1).

Although necessarily not occurring with constituents expressing the shadow argument (see (58)a), the verb *vacilar* (≅ tremble due to structural instability) occurs with expressions referring the cause of the instability, as the examples in (58)b illustrate.

(58) a. #O cais vacilou por causa de/devido a/com a instabilidade.
(The pier trembled because of/due to/ with instability)

b. O cais vacilou por causa de/devido a/com o choque/a violência das ondas.
(The pier trembled because of/due to/with the impact/violence of the waves)

6.2.3 Observed regularities

Although the goal of the analysis presented so far is not necessarily conditioned or directed by the need to establish patterns, the modeling of Portuguese verbs of movement in GL, focusing on the lexicalization of the semantic components considered – SOURCE, GOAL, DIRECTION, PATH, FIGURE, GROUND, INTENTION and CAUSE –, allows us to draw some generalizations.

With the exception of FIGURE, all components are lexicalized in the set of verbs observed:

- GOAL: {enjaular}$_V$ (≅ cage), {deitar-se}$_V$ (≅ lie down), for instance;
- SOURCE: {desencaixotar}$_V$ (≅ unbox), for instance;
- DIRECTION: {recuar}$_V$ (≅ move back), {subir 2}$_V$ (≅ go to on top of), for instance;
- PATH: {contornar}$_V$ (≅ circumvent), {circum-navegar}$_V$ (≅ circumnavigate), for instance;
- MANNER: {cambalear}$_V$ (≅ walk unsteadly), {estremecer}$_V$ (≅ tremble suddenly and occasionally), for instance;
- GROUND: {oscilar}$_V$ (≅ oscilate), {voar}$_V$ (≅ fly), for instance;
- INTENTION: {embalar}$_V$ (≅ rock), {fugir}$_V$ (≅ run away), for instance;
- CAUSE: {vacilar}$_V$ (≅ tremble due to structural instability).

Lexicalization corresponds necessarily to shadow arguments, obtained by the shadowing of true arguments of the hyperonym, as for instance in the case of {enjaular}$_V$ (≅ put inside of a cage; cage) or by the addition of new shadow arguments to the hyponym argument structure, as observable in the verb {subir 1}$_V$ (≅ move up/upwards).

However, and as expected, the meaning specifications denoted by hyponym verbs is always reflected at the argument structure level, whether corresponding to shadow arguments, whether corresponding only to the incorporation of semantic restrictions. Also, in some cases, these are also reflected in event and qualia structures.

The following regular properties, directly concerning the specification of the particular semantic components considered, were observed in the set of verbs analyzed:

i) The specification of GOAL and SOURCE occurs in transition denoting verbs and results in Aktionsart shifts with respect to the hyperonyms, denoting activity type events. The verbs that specify GOAL and SOURCE denote accomplishment or achievement type events. GOAL and SOURCE denoting arguments are expressed by prepositional phrases in change of location verbs and by prepositional or adverbial phrases in change of position verbs.

ii) DIRECTION is lexicalized in the semantic content of the hyponym and may contribute to Aktionsart shifting.

iii) The lexicalization of PATH requires also the presence of a new true argument, syntactically realized by a NP, replacing the default arguments denoting SOURCE and GOAL. In the case of incorporation of restrictions on PATH, the new true argument is syntactically expressed by a PP introduced by the preposition *por* (through) and the event denoted by the verb is typically an activity type event.

iv) The lexicalization of MANNER corresponds to new shadow arguments in the argument structure of the verb. These can correspond to concepts lexicalized by adverbial expressions or by prepositional phrases denoting instrument.

v) FIGURE is never lexicalized. The incorporation of restrictions on FIGURE is reflected in selection constraints on the arguments (second true argument in the case of dyadic verbs; first true argument in the case of monadic verbs).

vi) GROUND lexicalization expresses reference location or environment, or refers to specific properties of the FIGURE or of the PATH directly related to the motion event. Also, the specification of GROUND can be related to the definition of the PATH of the movement event. In this case, GROUND arguments are syntactically realized by NPs in object position.

vii) INTENTION is always lexicalized and results in a new shadow argument, corresponding to a prepositional phrase, introduced by the prepositional expression *para/com a intenção de* (≅ to/for, with the intention of), that selects an event complement. INTENTION lexicalization is also reflected in the qualia structure of the verbal items, as value of the Telic role.

viii) The one case of lexicalization of CAUSE resulted in the increase of the number of shadow arguments of the argument structure of the verb, with respect to the number of arguments selected by its hyperonym, being also reflected in the qualia structure as value of the Agentive role, since it refers to the causal chain or origin of the event.

The modeling of these verbs in GL framework allows us to represent them with regard to semantic and syntactic properties such as argument number and kind and Aktionsart properties. The integration of the lexical items in wordnet allows us the access to inherited information as well as to other lexical structures that provide information regarding subcategorization properties and qualia structure values.

6.3 Co-hyponyms compatibility

One of the immediate consequences of modeling the semantic content of lexical items using rich informational structures is the possibility of relating co-occurrence restrictions to semantic properties. The association of qualia information, for instance, to lexicon models is nowadays a commonly accepted strategy to enhance the expressive power of the models, namely aiming at the construction of lexica for computational purposes[2].

Mendes & Chaves (2001) consider that the association of qualia information to synsets enables the WordNet model to account for issues such as co-hyponyms compatibilities. By recognizing the different aspects of the meaning specification involved in the hyperonymy/hyponymy relation through the qualia values of each synset, it is possible to predict the compatibility of co-hyponyms, as briefly discussed in chapter 3, section 3.2.4.1 and illustrated with the example repeated here:

(59) a. This dog is a pit-bull [formal] and a police-dog [telic].

b. *This dog is a pit-bull [formal] and a Saint-Bernard [formal].

This contrast refutes the assumption that all co-hyponyms, as specifications of the same base concept, are incompatible, demonstrating that this incompatibility depends on the type of specific information each hyponym denotes.

To account for this phenomenon in nominal items, Mendes & Chaves (2001) assume a unification operation. Incompatibility is expressed considering that co-hyponym nouns are incompatible if their qualia structures do not unify:

(60) Two qualia structures do not unify if there is a role Q from two nominal qualia structures [Q=R1] and [Q=R2] where values R1 and R2 exist such that $\neg(R1 = R2) \wedge (\neg subsumes(R1, R2) \wedge \neg subsumes(R2, R1))$.

In other words, two co-hyponyms are incompatible if the values for a same qualia role in their qualia structures are not equal and if one of the values is not subsumed by the other.

The second conjunct in the last restriction $(\neg subsumes(R1, R2) \wedge \neg subsumes(R2, R1))$ assures that co-hyponyms with different values for the same qualia role are compatible when one of these values is a subtype of the other, such as in the cases in (61).

[2] See Fabre & Sébillot (1999), Busa et al. (2001), De Boni & Manandhar (2002), O'Hara & Wiebe (2003a), Veale (2003), among others.

(61) a. This house appliance is a fridge and a deep freezer.

b. fridge

$$\begin{bmatrix} \text{ARGSTR} = [\text{P-ARG1} = \text{x} : [\text{house appliance}]] \\ \text{QUALIA} = \begin{bmatrix} \text{TELIC} = \text{preserve}(\text{e}1, \text{x}, \text{y} : \text{food}) \\ \ldots \end{bmatrix} \end{bmatrix}$$

c. deep freezer

$$\begin{bmatrix} \text{ARGSTR} = [\text{P-ARG1} = \text{x} : [\text{house appliance}]] \\ \text{QUALIA} = \begin{bmatrix} \text{TELIC} = \text{freeze}(\text{e}1, \text{x}, \text{y} : \text{food}) \\ \ldots \end{bmatrix} \end{bmatrix}$$

c. $\{\text{freeze}\}_V$ is hyponym of $\{\text{preserve}\}_V$

The values R1 and of R2, *preserve(*e_1*, x, y:food)* and *freeze(*e_1*, x, y:food)*, of the Telic role are not equal, but *freeze* is a subtype *preserve*, accounting for the compatibility of these co-hyponyms. Also, and given the inheritance allowed for in wordnets, indirect co-hyponyms (sharing a common hyperonym high in the structure) are covered by this procedure. As pointed out by the authors, it should be possible to extend this operation to any item in the lexicon.

Let us, then, consider the Portuguese verbs of movement case, in (62):

(62) a. Ele saiu da sala, correndo.
(He exited from the room, running)

b. *Ele saiu da sala, entrando.
(He exited from the room, entering)

c. *Ele correu da sala, andando.
(He ran from the room, walking)

These examples demonstrate that compatibility between verbal co-hyponyms is also dependent on the semantic content of each verb, predicting that the qualia unification operation can also account for co-occurrence restrictions of co-hyponyms verbs. In the examples presented here, it seems that verbs that lexicalize the same semantic component cannot co-occur, (62)b and c, whereas verbs lexicalizing different semantic components are compatible, see (62)a. However, verbs such as *sair* (≅ move out) and *recuar* (≅ move back) that lexicalize DIRECTION are not incompatible, i.e. they can co-occur:

(63) a. Ele saiu da sala, recuando.
(He move up, moving back)

As already demonstrated in the previous part of this work, verbal qualia structure is somewhat different from that of nouns, being qualia unification as stated by Mendes & Chaves (2001) not directly applicable:

(64) a. {sair}$_V$ (exit)

$$\begin{bmatrix} \ldots \\ \text{QUALIA} = \begin{bmatrix} \text{AGENTIVE} = \text{exit}(e_1, x, t, u) \\ \text{FORMAL} = \text{exit\_state}(e_2, x, w) \end{bmatrix} \end{bmatrix}$$

b. {entrar}$_V$ (enter)

$$\begin{bmatrix} \ldots \\ \text{QUALIA} = \begin{bmatrix} \text{AGENTIVE} = \text{enter}(e_1, x, t, v) \\ \text{FORMAL} = \text{enter\_state}(e_2, x, w) \end{bmatrix} \end{bmatrix}$$

c. {correr}$_V$ (run)

$$\begin{bmatrix} \ldots \\ \text{QUALIA} = \left[\text{AGENTIVE} = \text{run}(e_1, x, z)\right] \end{bmatrix}$$

d. {andar}$_V$ (walk)

$$\begin{bmatrix} \ldots \\ \text{QUALIA} = \left[\text{AGENTIVE} = \text{walk}(e_1, x, z)\right] \end{bmatrix}$$

e. {recuar}$_V$ (move back)

$$\begin{bmatrix} \ldots \\ \text{QUALIA} = \left[\text{AGENTIVE} = \text{move\_back}(e_1, x, t)\right] \end{bmatrix}$$

Given the conditions established for the qualia unification operation presented above, none of these verbs are compatible, since they all present different values for the same qualia roles and none of these values subsumes the other. In fact, the validation of the subsumption is, on its turn, completely redundant, since the values in the qualia structure refer to semantic predicates that are subtypes of the same predicate, i.e., *move*.

To assume the unification operation as directly applying to verb qualia structures is to predict that all these verbs are incompatible, when in fact they are not, not accounting, at the same time, for the fact that the verbs *correr* (run) and *andar* (walk) are in fact incompatible. This seems to show that qualia structure unification is not directly applicable to the verbal lexicon.

Nonetheless, the data presented regarding compatibility among co-hyponyms, and the fact that it is possible to relate this phenomenon with the semantic content of verbs, constitute evidence that qualia unification is occurring at some level.

6.3.1 Indirect qualia unification

As explored in chapter 4, verbal qualia structure is fulfilled with semantic predicates expressing the relations between the arguments of a verb: Formal role values are semantic predicates corresponding to actual state of affairs, typically states; Agentive role values are semantic predicates involved in the causal chain of complex events, typically processes; Telic role values are semantic predicates that express the specific function of the denoted event; and

Constitutive role values are semantic predicates that form the event whose repetition constitutes a process.

However, it is at the argument structure level that the logical arguments of a predicate are listed. As we have demonstrated so far, verbal argument structure necessarily reflects the specification of the semantic elements responsible for the meaning specificities that distinguish hyperonyms from their hyponyms, and co-hyponyms from each other. For this reason, it is possible to assume that verbal co-hyponym compatibility is regulated by indirect qualia unification, proposed in the following terms:

(65) Two co-hyponym verbs are incompatible iff the arguments in their argument structures refer to incompatible co-hyponyms, i.e. if the qualia structures of these arguments do not unify.

Let us reconsider the verbs in (64). Indirect qualia structure unification enables us to predict that the co-hyponym verbs *sair* (exit) and *entrar* (enter) (besides being antonyms) and *correr* (run) and *andar* (walk) are incompatible given that the qualia structure of their shadow arguments, referring to co-hyponyms nodes, do not unify. Taking as example the AVMs in (66) and (67), it is possible to verify that: proper arguments (P-ARG$_1$) have the same value, the same happening with the first argument of these verbs (necessary conditions for these verbs to co-occur in contexts where incompatibility is tested); shadow arguments values are co-hyponyms (as attested by the hyperonym relation established at the proper-argument level of the structures), but are not compatible co-hyponyms, since their qualia structures do not unify; and finally, default arguments have also equal values. Naturally, only non-inherited arguments are at stake, here.

(66) {andar} (≅ move using limbs)

$$\begin{bmatrix} \text{ARGSTR} = \begin{bmatrix} \text{P-ARG}_1 = \text{e} : [\{\text{mover-se}\}_\text{V}\ (\text{move})] \\ \text{ARG}_1 = \text{x} = [\{\text{animal}\}_\text{N}\ (\text{animal})] \\ \text{S-ARG}_1 = \text{t} = \begin{bmatrix} [\text{using limbs}] \\ \text{ARGSTR} = \begin{bmatrix} \text{P-ARG}_1 = \text{m} : [\text{manner}] \\ \text{ARG}_1 = \boxed{1} \end{bmatrix} \\ \text{QUALIA} = [\text{FORMAL} = \text{using\_limbs}(\text{r}, \boxed{1})] \end{bmatrix} \\ \text{D-ARG}_1 = \text{u} = [\{\text{de}\}_\text{Prep}\ (\text{from})\ldots] \\ \text{D-ARG}_2 = \text{t} = [\{\text{para}\}_\text{Prep}\ (\text{to})\ldots] \\ \end{bmatrix} \\ \ldots \end{bmatrix}$$

(67) {correr} (≅ move fast using limbs and projecting the body forward)

$$\left[\begin{array}{l} \text{ARGSTR} = \left[\begin{array}{l} \text{P-ARG}_1 = e: [\{\text{mover-se}\}_V\ (\text{move})] \\ \text{ARG}_1 = x = [\{\text{animal}\}_N\ (\text{animal})] \\ \text{S-ARG}_1 = t = \left[\begin{array}{l} [\text{fast, using limbs, and projecting the body forward}] \\ \text{ARGSTR} = \left[\begin{array}{l}\text{P-ARG}_1 = m: [\text{manner}] \\ \text{ARG}_1 = \boxed{1}\end{array}\right] \\ \text{QUALIA} = [\text{FORMAL} = \text{fast\_using\_limbs}(r, \boxed{1})]\end{array}\right] \\ \text{D-ARG}_1 = u = [\{\text{de}\}_{Prep}\ (\text{from})\ldots] \\ \text{D-ARG}_2 = t = [\{\text{para}\}_{Prep}\ (\text{to})\ldots] \end{array}\right] \\ \ldots \end{array}\right]$$

Conversely, the co-hyponym verbs *sair* (exit) and *correr* (run) are compatible since there are no co-hyponym arguments in their structure: proper and first arguments have equal values; shadow arguments values are not co-hyponyms (as attested by the hyperonym relation established at their respective proper argument level), and the second argument of the verb *sair* (exit) is not present in the argument structure of the verb *correr* (run), nor it is co-hyponym of the default argument of this verb, the same occurring with the second argument of the verb *sair* (exit).

(68) $\{\text{sair}\}_V$ (≅ move out; exit)

$$\left[\begin{array}{l} \text{ARGSTR} = \left[\begin{array}{l} \text{P-ARG}_1 = e: [\{\text{mover}-\text{se}\}_V\ (\text{move})] \\ \text{ARG}_1 = \boxed{1}\ x = [\{\text{entidade}\}_N\ (\text{entity})] \\ \text{S-ARG}_1 = t = \left[\begin{array}{l} \{\text{para}\}_{Prep}\ (\text{to, in the direction of}) \\ \text{ARGSTR} = \left[\begin{array}{l}\text{P-ARG}_1 = r \\ \text{ARG}_1 = y = \{\text{fora}\}_N\ (\text{out})\end{array}\right] \\ \text{QUALIA} = [\text{FORMAL} = \text{to}(r, \boxed{2}, y)]\end{array}\right] \\ \text{ARG}_2 = w = \left[\begin{array}{l} \{\text{de}\}_{Prep}\ (\text{from}) \\ \text{ARGSTR} = \left[\begin{array}{l}\text{P-ARG}_1 = r \\ \text{ARG}_1 = y = [\{\text{entidade concreta}\}_N\ (\text{physical entity})]\end{array}\right] \\ \text{QUALIA} = [\text{FORMAL} = \text{from}(r, \boxed{1}, y), \neg\text{at}(e_1, \boxed{1}, y)]\end{array}\right] \end{array}\right] \\ \ldots \end{array}\right]$$

The slight reformulation of the qualia unification procedure proposed by Mendes & Chaves (2001) enables to accounting also for verbal co-hyponyms incompatibility.

To account for co-hyponyms compatibilities, besides being useful in what concerns inference devices "since negative propositions are harder to prove without resorting to a *closed world assumption* (if it isn't provable in the present model, then it is considered false)" (Mendes & Chaves 2001:111), motivates, once more, the use and the potential of rich informational structures for modeling lexical-semantic properties of the lexical items.

6.4 Conclusions

This chapter was dedicated to the deeper analysis of Portuguese verbs of movement and to the modeling of the information in their lexical entries, in GL model, and considering the organization of the lexical items in a wordnet.

Following the formulations proposed in chapter 4, we achieved a representation of these verbs that accounts for lexical inheritance structure and percolation of information within the lexicon, the reflections of semantic incorporation in the semantic and syntactic properties of hyponym verbs and verbal co-hyponyms compatibility.

On the one hand, the recursive use of available lexical structures, corresponding to the hyponym relation with other nodes in the network, established as values of the proper arguments, allows the percolation of information through the hyponymy trees, enabling a coherent and economic codification of the information. Also, the resulting lexical structures demonstrate that the hyponymy relation is capable of replacing a semantic type lattice in what concerns establishing and defining semantic properties by subtyping strategies. The use of available lexical structures also allows the direct and economic description of significant subcategorization properties.

On the other, the permeability granted by the GL model principles, in particular underspecification and co-composition, assures the necessary flexibility to explain the diversity of syntactic behaviors of the lexical items, directly related to lexical semantics properties.

The next logical task is to address the integration of GL structures in wordnet model, the subject of the next chapter.

7. GL structures in WordNet.PT

7.0 Introduction

Our goal to contribute to the definition of a computational lexicon that models the semantic and syntactic properties of the lexical items, as discussed earlier, is achieved through the integration of informational structures in a hierarchical lexicon. These reasons motivated our choice of the frameworks adopted here: the Generative Lexicon (GL), that provides the structured lexical entries, and WordNet (WN) that, by its nature, provides the necessary lexical hierarchy that allows the access to other structures in the lexicon. This strategy of modeling is possible since, as it will be demonstrate in this chapter, the integration of additional information in the WN model does not compromise its architecture.

The association of semantic and syntactic information to WN model has been, since its appearance, the subject matter of several researches (see Kohl et al. (1998), Leacock & Chodorow (1998), Harabagiu & Moldovan (1998), Agirre & Martinez (2002), for instance). On the one hand, this might show the insufficiency of the information conveyed by the lexical-conceptual relations established in this first model for many language processing tasks, but indicates, on the other, the variety of applications for which such a lexicon model is useful and its permeability to enhancements.

As discussed in chapter 2, one of the major efforts of enhancement of the WordNet model resulted in EuroWordNet (EWN), reflecting research on lexical semantics, lexicon organization and on WordNet as a lexical resource for a variety of applications. If we consider the new relations implemented in EWN, presented in (1), it is possible to identify the major concerns that originated them, namely the need for more comprehensive lexical-conceptual relations ((1)a, b and c), argument selection relations ((1)d, e and f) and cross part of speech relations ((1)g and h).

(1) Set of relations of EWN that do not existed in WordNet

a. NEAR SYNONYM relation
b. XPOS NEAR SYNONYM relation

c. NEAR ANTONYM relation
d. ROLE/INVOLVED relation
e. CO_ROLE relation
f. IN MANNER/MANNER OF relation
g. INSTANCE OF/HAS INSTANCE relation
h. BE IN STATE/STATE OF relation

Also WordNet.PT (WN.PT), build up within EWN model for Portuguese, has since its beginning provided grounds for fundamental research on Lexical Semantics[1], profiting, at the same time, from the research developed, as it is reflected by the set of relations established in this particular model (see chapter 2, section 2.1.3). Our goal here is, thus, consistent with the tendency adopted for developing WN.PT and follows the dynamics of the project.

In this chapter, we present our proposal for integrating GL structures in WN.PT, a process suitable for any wordnet, considering the analysis presented so far and focusing on the results for Portuguese verbs of movement. Note, however, that the modeling strategies here proposed are transversal to all POS, covering, thus, the entire lexicon.

The first section of this chapter concerns the integration of argument structure in WN.PT, and the proposal for implementing new relations. Section 2 addresses the specification of qualia structure in WN.PT using mostly already established relations. Section 3 proposes the integration of event structures in WN.PT, considering the particular specificities of this representation level. Finally, sections 4 and 5 present, respectively, the resulting set of relations available in WN.PT and the conclusions of this chapter.

7.1 Integrating argument structure

The integration of selection properties in WordNet model has been proposed, for instance, in Agirre & Martinez (2002) through the extension of previous statistical models of word-to-class to class-to-class preferences that allows the learning of selectional preferences for classes of verbs, integrating its results in WordNet 1.5. The authors assume that different senses of a given verb may display different selectional preferences and that classes of verbs may share preferences, and associate to each sense of a given word (synset) statistic information on selectional preferences.

[1] See Marrafa & Mendes (2007, 2006), Mendes (2007, 2006), Amaro *et al.* (2006), Amaro (2006, 2005), Marrafa *et al.* (2006, 2005, 2004a, 2004b), Marrafa (2005, 2004, 2003a, 2003b, 2002), Ribeiro et al. (2004a, 2004b), Chaves (2001) and Mendes & Chaves (2001).

The EWN model, on its turn, describes selectional properties through role relations, in the sense that these establish relations between event denoting nodes and the nodes denoting the participants *typically* involved in those events.

(2) a. ROLE AGENT/INVOLVED AGENT relation
a N_1 is the one/that who/which does V_1/N_2, typically intentionally

b. ROLE PATIENT/INVOLVED PATIENT relation
a N_1 is the one/that who/which undergoes V_1/N_2

c. ROLE RESULT/INVOLVED RESULT relation
a N_1 comes into existence as result of/is the result of/is created by V_1/N_2

d. ROLE INSTRUMENT/INVOLVED INSTRUMENT relation
a N_1 is the instrument/what is used to V_1/N_2

e. ROLE LOCATION/INVOLVED LOCATION relation
a N_1 is the place where V_1/N_2 happens

f. ROLE SOURCE DIRECTION/INVOLVED SOURCE DIRECTION relation
a N_1 is the place from where V_1ing/N_2 begins/starts/happens or one V_1s

g. ROLE TARGET DIRECTION/INVOLVED TARGET DIRECTION relation
a N_1 is the place to where V_1ing/N_2 happens or one V_1s

h. ROLE /INVOLVED relation
a N_1 is the one/that who/which is typically involved in V_1ing/N_2

Role relations are established between verbal synsets and deverbal nouns denoting events synsets and nominal synsets. Role relations are based on thematic role assignment, thus assuming some form of syntactic mapping function: agents are commonly assumed to be syntactically realized in subject position, patients are commonly assumed to be syntactically realized in object position, instruments are commonly assumed to be syntactically realized in oblique position and so forth, as the EWN tests presented above for each of the role relations available show. Note however, that these relations are established to connect nodes that lexicalize a given thematic function of a given event or whose thematic function is necessarily or typically associated to the concept denoted:

(3) a. $\{teacher\}_N$ ROLE AGENT $\{teach\}_V$/$\{teach\}_V$ INVOLVED AGENT $\{teacher\}_N$
a *teacher* is the one/that who/which *teaches*, typically intentionally

b. $\{pupil\}_N$ ROLE PATIENT $\{teach\}_V$/$\{teach\}_V$ INVOLVED PATIENT $\{pupil\}_N$
a *pupil* is the one who undergoes *teaching*

c. $\{building\}_N$ ROLE RESULT $\{build\}_V$/$\{build\}_V$ INVOLVED RESULT $\{building\}_N$
a *building* comes into existence as result of/is the result of/is created by *building*

d. {scissors}$_N$ ROLE INSTRUMENT {cut}$_V$/{cut}$_V$ INVOLVED INSTRUMENT {scissors}$_N$
scissors is the instrument/what is used to *cut*

e. {pool}$_N$ ROLE LOCATION {swim}$_V$/{swim}$_V$ INVOLVED LOCATION {pool}$_N$
a *pool* is the place where *swimming* happens

f. {homeland}$_N$ ROLE SOURCE DIRECTION {emigrate}$_V$/{emigrate}$_V$ INVOLVED SOURCE DIRECTION {homeland}$_N$
a *homeland* is the place from where one *emigrates*

g. {prison}$_N$ ROLE TARGET DIRECTION {imprison, incarcerate}$_V$/{imprison, incarcerate}$_V$ INVOLVED TARGET DIRECTION {prison}$_N$
a *prison* is the place to where one *imprisons/incarcerates*

h. {vehicle}$_N$ ROLE {travel}$_V$/{vehicle}$_N$ INVOLVED {travel}$_V$
a *vehicle* is that which is typically involved in *traveling*

Role relations, by their intrinsic characteristics, do not explicitly express subcategorization properties. As stated before, these relations express the association between events and the participants *typically* involved in them. Considering the characteristics and goals of lexical-conceptual models, these relations, as well as all the other relations considered, are designed to represent the semantic and conceptual properties of the concepts lexicalized in a given language.

GL argument structure, on the other hand, consists in a level of representation in which the number and type of arguments of a lexical item are stated, contemplating the definition of the semantic properties of its logical arguments, as well as syntactic mapping information. For this reason, we feel that the integration of argument information in WN.PT, and the use of the available structures in the network, results in the increase of the relevant information on the semantic and syntactic properties of the lexical items, contributing to their description and enhancing the use of wordnets as resources for meaning computation purposes.

Our proposal is to express argument structure in WN.PT, including default and shadow arguments, through three new relations and a new order feature:

i) **SELECTS/IS SELECTED BY** relation, for true arguments;

ii) **INCORPORATES /IS INCORPORATED IN** relation, for shadow arguments;

iii) **SELECTS BY DEFAULT /IS SELECTED BY DEFAULT** relation, for default arguments;

iv) *order* feature, with numerical tags, for establishing the order of arguments, which, as defined in GL, express the constraints on the syntactic mapping: arguments are represented in a list from the less oblique one to the more oblique one[2].

The new relations are defined as follows:

(4) SELECTS/ IS SELECTED BY relation

$\{synset\}_1$	SELECTS	$\{synset\}_2$ and
$\{synset\}_2$	IS SELECTED BY	$\{synset\}_1$

iff:

i) $x \in \{synset\}_1$ and $y \in \{synset\}_2$, and the syntactic realization of x requires the syntactic realization of y, or of z, $z \in \{synset\}_3$ hyponym of $\{synset\}_2$.

(5) INCORPORATES/IS INCORPORATED IN relation

$\{synset\}_1$	INCORPORATES	$\{synset\}_2$ and
$\{synset\}_2$	IS INCORPORATED IN	$\{synset\}_1$

iff:

i) the concept denoted by the $\{synset\}_1$ entails the specific concept lexicalized by the $\{synset\}_2$;
ii) $x \in \{synset\}_1$ and $y \in \{synset\}_2$, and the co-occurrence of x and y is only licensed by subtyping or specification processes; and
iii) in case of conjoint incorporations, ii) only applies to the element with reference potential.

(6) SELECTS BY DEFAULT/IS SELECTED BY DEFAULT BY relation

$\{synset\}_1$	SELECTS BY DEFAULT	$\{synset\}_2$ and
$\{synset\}_2$	IS SELECTED BY DEFAULT	$\{synset\}_1$

iff:

i) the concept denoted by the $\{synset\}_1$ entails but not necessarily the concept denoted by the $\{synset\}_2$;
ii) $x \in \{synset\}_1$ and $y \in \{synset\}_2$ and the co-occurrence of x and y is only licensed by subtyping or specification processes; and
iii) in case of conjoint default selections, ii) only applies to the element with reference potential.

[2] Order here refers to the so called basic order of constituents. The list of arguments provides their basic position, not accounting for other possible syntactic positions.

The *order* feature is used to indexate the arguments established by SELECTS, INCORPORATES and SELECTS BY DEFAULT relations to a given order, established by an argument list, <1, 2, …, n>, determined at the event structure level (presented in section 7.3 ahead). *Order* feature values are expressed by natural numbers.

To exemplify the use of these new relations, let us consider for instance, the argument structures of the nodes {pôr, colocar}$_V$ (≅ put), {meter}$_V$ (≅ put inside of) and {enjaular}$_V$ (≅ cage) whose simplified argument structures are here in (7) a, b and c, respectively.

(7) a. {pôr, colocar}$_V$ (≅ put)

$$\left[\begin{array}{ll} \text{ARGSTR} = & \left[\begin{array}{l} \text{P-ARG}_1 = e : [\{\text{mover, deslocar}\}_V \text{ (move)}] \\ \text{ARG}_1 = x = [\{\text{entidade}\}_N \text{ (entity)}] \\ \text{ARG}_2 = y = [\{\text{objecto}\}_N \text{ (object)}] \\ \text{D-ARG}_1 = u = [\{\text{de}\}_{Prep} \text{ (from)}] \\ \text{S-ARG}_1 = v = [\{\text{para}\}_{Prep} \text{ (to)}] \\ \text{ARG}_3 = w = [\{\text{em}\}_{Prep} \text{ (in/at)}] \end{array}\right] \\ \ldots & \end{array}\right]$$

b. {meter}$_V$ (≅ put inside of)

$$\left[\begin{array}{ll} \text{ARGSTR} = & \left[\begin{array}{l} \text{P-ARG}_1 = e : [\{\text{pôr, colocar}\}_V \text{ (put)}] \\ \text{ARG}_1 = x = [\{\text{entidade}\}_N \text{ (entity)}] \\ \text{ARG}_2 = y = [\{\text{objecto}\}_N \text{ (object)}] \\ \text{D-ARG}_1 = u = [\{\text{de}\}_{Prep} \text{ (from)}] \\ \text{S-ARG}_1 = v = [\{\text{para dentro de}\}_{Prep} \text{ (to inside)}] \\ \text{ARG}_3 = w = [\{\text{em}\}_{Prep} \text{ (in/at)}] \end{array}\right] \\ \ldots & \end{array}\right]$$

c. {enjaular}$_V$ (≅ cage)

$$\left[\begin{array}{ll} \text{ARGSTR} = & \left[\begin{array}{l} \text{P-ARG}_1 = e : [\{\text{meter}\}_V \text{ (put)}] \\ \text{ARG}_1 = x = [\{\text{entidade}\}_N \text{ (entity)}] \\ \text{ARG}_2 = y = [\{\text{animal}\}_N \text{ (animal)}] \\ \text{D-ARG}_1 = u = [\{\text{de}\}_{Prep} \text{ (from)}] \\ \text{S-ARG}_1 = w = \left[\begin{array}{ll} \{\text{em}\}_{Prep} \text{ (in/at)} & \\ \text{ARGSTR} = & \left[\begin{array}{l} \text{P-ARG}_1 = e \\ \text{ARG}_1 = y = \left[\begin{array}{l} \{\text{jaula}\}N \text{ (cage)} \\ \ldots \end{array}\right] \end{array}\right] \\ \ldots & \end{array}\right] \end{array}\right] \\ \ldots & \end{array}\right]$$

The following scheme illustrates how the new relations are used to express selection and subcategorization properties, differentiating the true, default and shadow arguments of a given lexical item and allowing the description of subcategorization properties.

(8) WN.PT fragment with new selection relations

{pôr, colocar}V [*mover* to a given location; ≃ *put*]
- HAS HYPERONYM → **{mover, deslocar}V** [≃ *move*]
- SELECTS → 1 **{entidade}N** [≃ entity]
- SELECTS → 2 **{objecto}N** [≃ object]
- SELECTS BY DEFAULT → 3 **{de}Prep** [indicator of source; ≃ from]
- INCORPORATES → 4 **{para}Prep** [indicator of goal; ≃ to]
- SELECTS → 5 **{em}Prep** [indicator of location; ≃ in/at]
- HAS HYPONYM
  - **{meter}V** [*pôr* inside of; ≃ *put inside of*]
    - SELECTS → 1 **{entidade}N** [≃ entity]
    - SELECTS → 2 **{objecto}N** [≃ object]
    - SELECTS BY DEFAULT → 3 **{de}Prep** [indicator of source; ≃ from]
    - INCORPORATES → 4 **{para dentro de}Prep** [≃ to inside of]
    - SELECTS → 5 **{em}Prep** [indicator of location; ≃ in/at]
    - HAS HYPONYM
      - **{enjaular}V** [*meter* in a cage; ≃ *cage*]
        - SELECTS → 1 **{entidade}N** [≃ entity]
        - SELECTS → 2 **{animal}N** [≃ animal]
        - SELECTS BY DEFAULT → 3 **{de}Prep** [indicator of source; ≃ from]
        - INCORPORATES → 4 **{em}Prep** [indicator of location; ≃ in/at] **conj 1:1**
        - INCORPORATES → 4 **{jaula}N** [≃ cage] **conj 2:1**

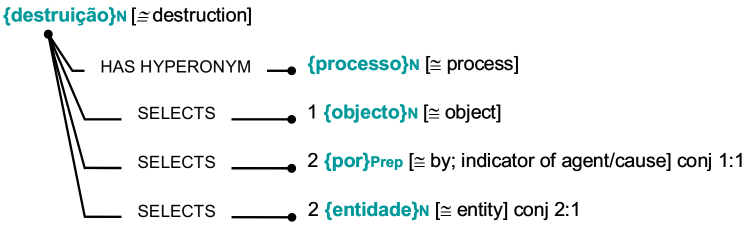
The combination of the SELECTS relation with the numerical tags of the order feature <1, 2, 3, ..., n> allows the extraction of the realization order of the arguments, according to their position in the list stated at event structure level.

The relation tagged with 1 indicates the argument that is realized in the less oblique position (in the case of verbs, in subject position, in the case of nouns, in object position mediated by a preposition), the relation tagged with 2 indicates the argument that is realized in object position, and so on.

The INCORPORATES relation, although referring to shadow arguments and thus typically not syntactically expressed, is also tagged to assure the extraction of the correct syntactic position in the case of syntactic realization, through subtyping or specification processes.

Order tags have to be associated to the argument list given at the synset level (as further explained ahead) so as to express the arguments position in the syntax. Features, however, convey additional information that is not visible to the current system, since it is not expressed through lexical-conceptual relations. Nonetheless, the systematic statement of this information associated to selection relations will provide the grounds to extract argument structure from relational resources such as wordnets.

(9) Example of the argument structure of a noun:



Interestingly, the integration of prepositions in the lexicon allows also to state selection properties in a more accurate way. For instance, the verbs *pôr, colocar* (≅ put) selects an argument that refers a given location. If we considered that this argument is of the semantic type *location* or that it denotes the concept of location, without considering prepositions in the lexicon, the strategy would be to state that this argument had to be of the semantic type `location`, or, in wordnets, to link the synset {*pôr, colocar*}$_V$ (≅ put) to the node that expressed the concept of location, in Portuguese lexicalized by the lexical items in {local, lugar}$_N$ (≅ location, place).

In both cases, the restriction to lexical items of the semantic type `location` or to the synset {local, lugar}$_N$ (≅ location, place) would result in an inaccurate representation of the selection restrictions established by the event denoted by these verbs: it is possible to put objects in the ground/street/garden/etc. – possible hyponyms of {location}$_N$ or lexical items characterized as being of the semantic type `location`, but it is also possible and frequent to put objects in tables/beds/closets/steps/etc., lexical items that denote objects that hardly would be characterized as having the semantic type `location` or as being hyponyms of the node {location}$_N$ in a wordnet. In fact, these examples show that the location meaning is achieved compositionally by the combination of the preposition and its complement meanings.

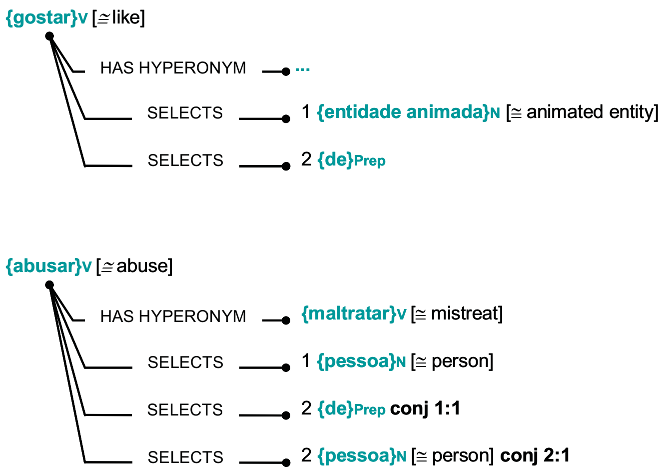
For these reasons, to indicate the prepositional synset {em}$_{Prep}$ (indicator of location; ≅ in/at) as selected by the verbs in {pôr, colocar}$_V$ (≅ put) results in a more accurate modeling of the selection and subcategorization properties of these verbs. Also, as established in chapter 5 and reflected in the SELECTS/IS SELECTED BY relation definition in (4) above, the determination of a given true argument considers the realization of the variants of the appointed node or of any of its hyponyms, describing in an economic fashion a multitude of possible syntactic realizations.

There is, however, a different issue to be addressed concerning the relations established between prepositional nodes and their selecting verbs, namely in the cases where the verb also imposes constraints on the complement of the preposition. The verb *enjaular* (≅ cage), above, illustrates this situation. In this case, the verb incorporates an argument referring to a specific location, a cage, introduced by the preposition *em*. Given that the wordnet model is a lexicon model, in which there are no nodes referring to phrases instead of lexical items (nor would it be desirable), it is necessary to devise a strategy to represent these cases.

Making use of the label conjunction available in EWN model (see chapter 2, section 2.1.2), we propose to model these cases by considering both the prepositional and the nominal nodes as selected by the verb (in the case of the verb *enjaular* (≅ cage)), constituting a shadow argument, thus the INCORPORATES relation, tagged with the same order tag, and directly linked to each other by the conjunction label. The EWN model allows for distinguishing the elements of a same conjunction by labeling them as conjunction 1, conjunction 2, and so on. This way, the relation established reads as follows: the synset {enjaular}$_V$ (≅ cage) INCORPORATES {em}$_{Prep}$ (≅ in/at) and INCORPORATES {jaula}$_N$ (≅ cage) that conjointly constitute the shadow argument of the event denoted by {enjaular}$_V$, being {em}$_{Prep}$ the first element of the conjunction 1 and {jaula}$_N$ the second element of the conjunction 1.

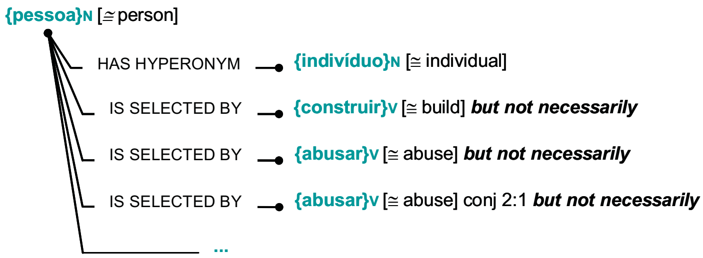
This strategy is also necessary for linking argument-marking prepositions to their selecting verbs, describing at the same time the selection properties, if relevant, of the verb. Let us consider, for instance, the verbs {gostar}$_V$ (≅ like) and {abusar}$_V$ (≅ abuse). Both subcategorize the argument-marking preposition *de*, while imposing different selection properties on the arguments (underspecified and persons, respectively). With this strategy it is possible to describe uniformly both cases (see (10)).

(10) WN.PT fragment



Default arguments – that is, arguments that refer to parameters required by the logical expression of the lexical item that can be syntactically expressed by subtyping or specification processes – in a first perspective, could be expressed using the SELECTS relation and the reversed label (that in EWN model marks by default the case where the inverse directions of a given relation are not symmetric, since all relations are bidirectional, see chapter 2, section 2.1.2), that can be paraphrased by the expression *not necessarily*. Note, however, that this option would result in a non coherent coding of default arguments, on the one hand, and of all the non symmetric relations, on the other, as exemplified in (11).

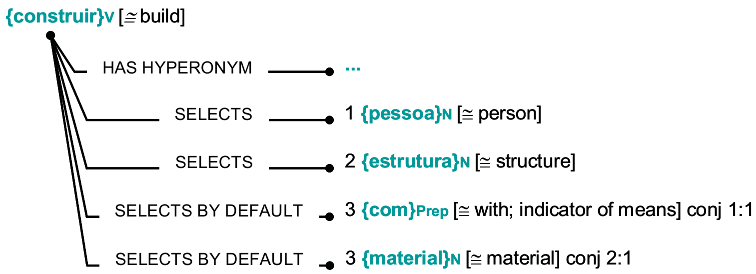
(11) Example of non-symmetric relations:



The non symmetric relation indicates that the node {pessoa}$_N$ (≅ person) is selected by {construir}$_V$ (≅ build) and {abusar}$_V$ (≅ abuse), but not necessarily, and not that {pessoa}$_N$ (≅ person) is selected by default by these verbs.

Also, default arguments, as shadow arguments, are only syntactically licensed by subtyping or specification processes, a property not shared with true arguments, established through the SELECTS/IS SELECTED BY relation. These issues led us to propose a new relation – SELECTS BY DEFAULT/IS SELECTED BY DEFAULT BY – defined in (6) above. The synset {construir}$_V$ (≅ build), in (12), illustrates the use of this relation.

(12) Example of SELECTS BY DEFAULT relation



The three new relations proposed here are, as demonstrated in this section, extendable to all POS in the lexicon. Selection relations can be used to characterize verbal, nominal, adjectival, adverbial and prepositional selection properties and subcategorization properties, since synsets are necessarily characterized regarding to the POS category.

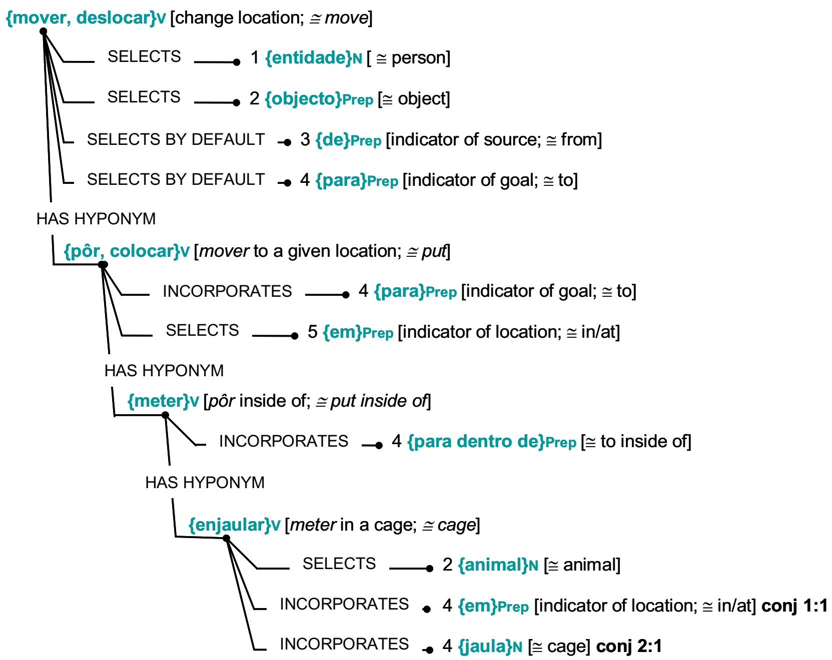
Note however, that this information alone is not sufficient to account for the syntactic realization of argument structure since there are general syntactic conditions that apply to the expression of arguments that do not constitute lexical syntactic properties, as for instance, the fact that nominal expressions do not take subject arguments, or that these noun arguments are always introduced by the preposition *de*. Nonetheless, the integration of argument structure information in wordnets results in the enrichment of the descriptive power of the model.

At this point, it is also necessary to discuss the coding options associated to the representation of this information, namely in what concerns information inheritance through the network, but also in what concerns decisions on the relevant and useful information to be stated in the network.

From the examples given above (see (8)), it is clear that the definition of the argument structure at each node of the hierarchy results, naturally, in the repetition of information. That is, hyponyms inherit at least part of the argument structure from their hyperonyms, but not necessarily all the argument structure information. The original wordnet model assumes that inheritance is monotonic (see Miller 1990), since the lexical-conceptual properties are necessarily inherited through hyponym relations: if $\{\text{car}\}_N$ is a subtype of $\{\text{vehicle}\}_N$, than $\{\text{car}\}_N$ has all the lexical-conceptual properties that define $\{\text{vehicle}\}_N$. However, the fact that argument structure information is not completely inherited by hyponym nodes motivates the assumption of lexical inheritance by default: if nothing is stated on the contrary, hyponyms inherit all the information that characterizes their hyperonyms. For this reason, it is only necessary to state the new and/or more specific information at the hyponym level. The inheritance device to operate over this resource must thus be able to restrict the cases in which the specific information denoted by the hyponym replaces or is added to the inherited information.

This way, the selection properties established at the hyponym level only reflect non-inherited information. The subtree for the nodes $\{\text{mover, deslocar}\}_V$ (≅ move), $\{\text{pôr, colocar}\}_V$ (≅ put), $\{\text{meter}\}_V$ (≅ put inside of) and $\{\text{enjaular}\}_V$ (≅ cage), here in (13), illustrates this case:

(13) Example of WN.PT fragment expressing non-inherited argument relations



The association of order feature values, which characterize the selection relations, to the list of arguments of a given lexical item results in relevant information for extracting argument structure information from wordnets, contributing also to the definition of the inheritance device.

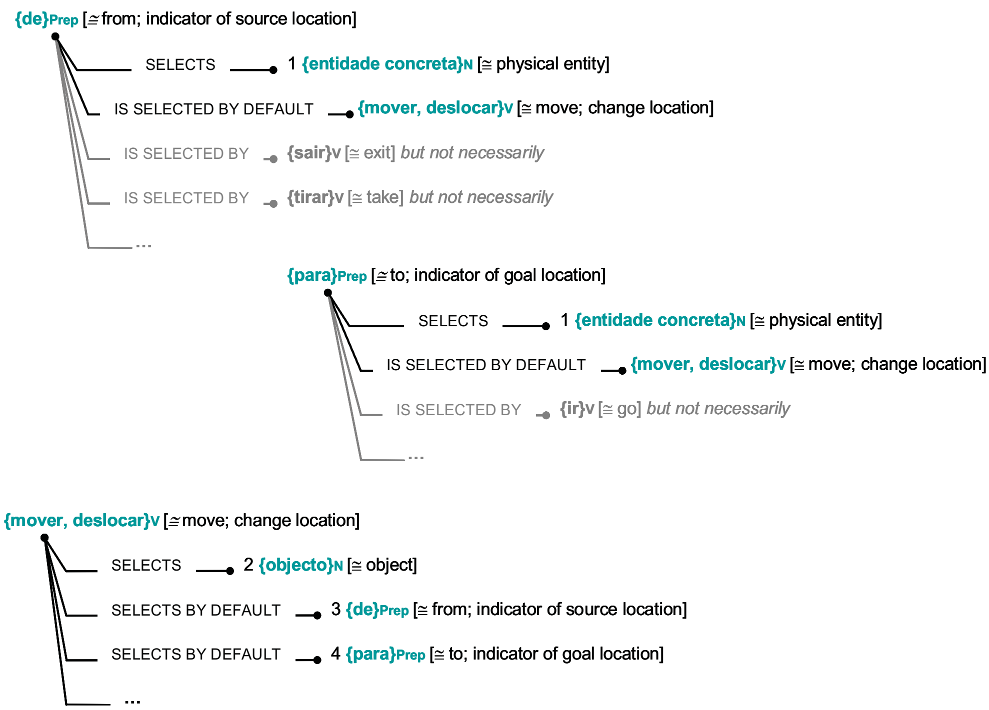
In what concerns the decision on the description of argument structure information, it is also necessary to discuss the benefits of overtly describing selection properties that do not correspond to specific restrictions over given concepts in the net (for instance, {entidade}$_N$ [≃ entity] as the first argument selected by the verbs in the synset {mover, deslocar}$_V$ [≃ change location], as opposed for instance, to the specific argument {jaula}$_N$ [≃ cage] incorporated in the verb {enjaular}$_V$ [≃ cage], or the prepositional node indicator of goal location selected by the verbs in the synset {pôr, colocar}$_V$ [≃ put]).

The option to state selection properties that do not correspond to specific restrictions over given concepts can result in non useful information, since general concepts denoting nodes (such as *entity*) are, more often than not, hyperonyms of concepts denoting nodes that by their specificities cannot be selected as arguments of a given lexical item. In these cases, the selection restrictions correspond to specific properties or *features* that a given argument must denote in order to be selected. The definition of a *feature* (or qualia, for instance) characterization system for postulating this type of restrictions is out of the scope of the present work. However, and as note for future research, we think that the enrichment of wordnets with selection and qualia information can constitute a path to reach this solution.

For the reasons presented here, the information on the selection properties of lexical items should also reflect distinctive properties of a given node of the network, and are considered as strict selection properties.

To conclude this section, let us observe the results of the expression of argument structure in a wordnet, by considering the information stated at the prepositional top nodes $\{de\}_{Prep}$ (indicator of source location; ≅ from) and $\{para\}_{Prep}$ (indicator of goal location; ≅ to) and at the top node of change of location verbs, the node $\{mover, deslocar\}_{V}$ (change location; ≅ move), for instance, here in (14):

(14) WN.PT fragment with selection relations



The relations established between these nodes express that the prepositions *de* and *para*, denoting the concepts of indicator of source and goal locations, respectively, occur with verbs of change of location, that correspond to the variants in the node $\{mover, deslocar\}_{V}$ or in any of its hyponyms. These prepositional nodes can also be selected by particular verbal nodes, such as $\{sair\}_{V}$ (≅ exit), $\{tirar\}_{V}$ (≅ take), $\{ir\}_{V}$ (≅ go), and so on, but not necessarily.

The top node {mover, deslocar}$_V$ (change location; ≅ move), on its turn is characterized by selecting as default arguments the nodes {de}$_{Prep}$ (≅ from) and {para}$_{Prep}$ (≅ to) and also by selecting objects – physical entities – as its second true argument.

7.2 Qualia specification

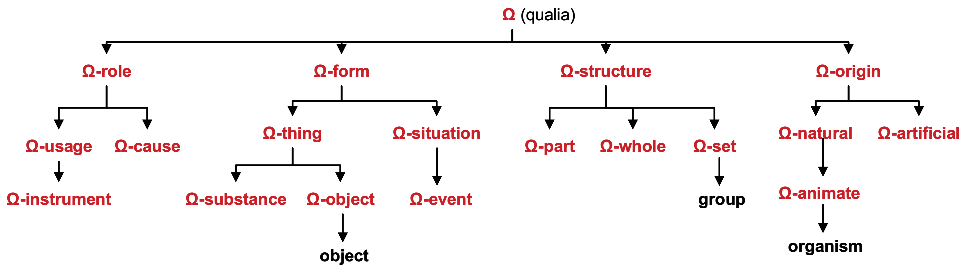
7.2.1 Qualia information and wordnets

As demonstrated in the previous chapters of this work, qualia information is necessarily related to the semantic content of the lexical items. Thus, its direct or indirect reference in lexical-conceptual models of the Lexicon is consistent with the nature of these models. WN structuring relation of hyperonymy/hyponymy, for instance, can be seen as referring to the Formal aspect of the semantic content of a given lexical item (see Pederson & Sørensen 2006), meronymy relations to the Constitutive aspect, cause relations to the Agentive aspect, and so on.

Qualia information in WN can also be expressed indirectly by additional information in the model, namely the glosses. Veale (2003), for instance, presents a proposal for automatically extracting agentive and telic properties from the sense glosses in WordNet 1.5 to account for metaphorical uses of lexical items, assuming that formal and constitutive properties are already established through hyperonymy and meronymy relations, respectively.

EWN model reflects to some extent the notion of qualia information, since GL qualia structure was used as the basis for its top-ontology, developed to provide a common framework for central concepts or synsets in wordnets for different languages (see Vossen (1999)). This resulted in establishing that each top-node expresses a qualia aspect: role (telic), form (formal), structure (constitutive) and origin (agentive), from which the rest of the net is constructed. The following figure depicts part of this ontology:

(15) Top-node ontology for EWN



(Excerpt from Vossen (2001: 8))

This top-node ontology reflects qualia aspects only at the taxonomy level, i.e., it does not reflect the multiple aspects of the meaning of the particular lexical items, as proposed by Pustejovsky (1995). Note, however, that the determination of the top-nodes for the lexicon is not an easy task given the general concepts denoted. Top-nodes, in spite of being underspecified, can be thought of as focusing on a given aspect or property, identifiable with qualia aspects. As observed by the author: "Rather than describing concepts as complex constructs that can be defined by multiple qualia, only one of them is selected, while different senses reflect separate dimensions of classification: functional, constitutional, or agentive hyperonyms." (Vossen 2001:5)

Based on EWN top-nodes ontology, Pederson & Sørensen (2006) propose to differentiate, in Danish wordnet (DanNet), taxonomic hyponyms (determined by hyperonymy and corresponding to the Formal qualia aspect) from orthogonal hyponyms, divided according to the other three qualia dimensions. This strategy results in the distinction of *constitutive* hyponyms (*idiot* 'idiot', *geni* 'genius', *smadrekasse* 'rattletrap', for instance), *telic* hyponyms (*vejtræ* 'roadside tree', *garvestof* 'tanning agent', *flugtbil* 'getaway car') and *agentive* hyponyms (*fodgænger* 'pedestrian' or *cyklist* 'cyclist') (examples from Pederson & Sørensen (2006: 5)). This approach, although accounting for co-hyponyms compatibilities, does not reflect the fact that, as referred before, individual lexical items can be characterized by multiple qualia, and that these properties are related to, for instance, selection restrictions or adjectival modification (see chapter 4, section 4.1.2.2).

In spite of the different goals and approaches, the research briefly discussed here reinforces the general need for more complete informational structures in the lexicon, on the one hand, but also shows that lexical-conceptual models convey, more or less directly (i.e. through already established relations or through auxiliary information such as glosses) a significant amount of relevant information regarding the internal semantic content of lexical items.

In the next section, we will present our proposal for integrating qualia information in WN.PT, using the already available relations in the model and its permeability to the implementation of new relations.

7.2.2 Qualia structure in WN.PT

The four qualia roles correspond to different aspects of the meaning of lexical items: the Constitutive role focuses on the relations between a given (semantic) object and its constituents or parts; the Formal role focuses on the stative properties that distinguish a given (semantic) object within its semantic domain; the Telic role concerns the information about the function or purpose of a given (semantic) object and the Agentive role focuses the origin or the causal chain involved in the bringing about of the (semantic) object. The values of these roles, in GL, are fulfilled with semantic predicates that express the relation between the different semantic objects that define the meaning of a lexical item.

In wordnets these relations can be expressed by lexical-conceptual relations established between the nodes of the net. As mentioned above, the meronym relation, for instance, establishes constitutive properties and the hyponym relation formal properties. Following this perspective, and having in mind that lexical-conceptual relations in wordnets reflect intrinsic or prototypical properties that characterize the concept lexicalized by each synset, we analyzed the relations available in WN.PT to determine which qualia properties, if any, these relations characterize.

The following table summarizes this analysis. WN.PT relations, in the second column, are grouped according to the qualia roles they refer to.

Qualia roles	WN.PT Relations	Examples
AGENTIVE	**RESULTS/ORIGINATES FROM**	{energia solar}$_N$ RESULTS/ORIGINATES FROM {sol}$_N$ {solar energy}$_N$ RESULTS/ORIGINATES FROM {sun}$_N$
	IS CAUSED BY (*IS CAUSED BY*)	{achar}$_V$ IS CAUSED BY {procurar}$_V$ {find}$_V$ IS CAUSED BY {search}$_V$
	IS THE RESULT OF (*ROLE RESULT*)	{assado}$_N$ IS THE RESULT OF {assar}$_V$ {roastbeef}$_N$ IS THE RESULT OF {roast}$_V$
	RESULTS FROM THE TRANSFORMATION OF (*CO RESULT PATIENT*)	{rã}$_N$ RESULTS FROM THE TRANSFORMATION OF {girino}$_N$ {frog}$_N$ RESULTS FROM THE TRANSFORMATION OF {tadpole}$_N$
	RESULTS FROM THE USE OF (*CO RESULT INSTRUMENT*)	{café}$_N$ RESULTS FROM THE USE OF {cafeteria}$_N$ {coffee}$_N$ RESULTS FROM THE USE OF {coffeepot}$_N$
	RESULTS FROM THE ACTION OF (*CO RESULT AGENT*)	{bread}$_N$ RESULTS FROM THE ACTION OF {padeiro}$_N$ {bread}$_N$ RESULTS FROM THE ACTION OF {baker}$_N$
CONSTITUTIVE	HAS AS PART (*HAS MERONYM*) HAS AS INDIVIDUATED PART (*HAS MERO PART*) HAS AS PORTION (*HAS MERO PORTION*) HAS AS MEMBER (*HAS MERO MEMBER*) HAS AS SUBSTANCE/MATERIAL (*HAS MERO MADEOF*) HAS AS SUBLOCATION (*HAS MERO LOCATION*)	{parede}$_N$ HAS AS PART {tijolo}$_N$ {wall}$_N$ HAS AS PART {brick}$_N$ {carro}$_N$ HAS AS INDIVIDUATED PART {roda}$_N$ {car}$_N$ HAS AS INDIVIDUATED PART {wheel}$_N$ {citrino}$_N$ HAS AS PORTION PART {gomo}$_N$ {citrus fruit}$_N$ HAS AS PORTION PART {[part of citris fruit] }$_N$ {club}$_N$ HAS AS MEMBER {member}$_N$ {member}$_N$ HAS AS MEMBER {club}$_N$ {café}$_N$ HAS AS SUBSTANCE/MATERIAL {cafeína}$_N$ {coffee}$_N$ HAS AS SUBSTANCE/MATERIAL {caffeine}$_N$ {parque de estacionamento}$_N$ HAS AS SUBLOCATION {lugar}$_N$ {parking lot}$_N$ HAS AS SUBLOCATION {parking space}$_N$
	HAS SUBEVENT (*HAS SUBEVENT*)	{respirar}$_V$ HAS SUBEVENT {inspirar}$_V$ {breathe}$_V$ HAS SUBEVENT {inhale}$_V$
	HAS TELIC SUBEVENT	{entristecer}$_V$ HAS TELIC SUBEVENT {triste}$_{Adj}$ {sadden}$_V$ HAS TELIC SUBEVENT {sad}$_{Adj}$

Qualia roles	WN.PT Relations	Examples
FORMAL	IS HYPONYM (SUBTYPE) OF (*has hyperonym*)	{animal}$_N$ IS HYPONYM (SUBTYPE) OF {ser vivo}$_N$ {animal}$_N$ IS HYPONYM (SUBTYPE) OF {living being}$_N$
	HAS AS A CHARACTERISTIC TO BE	{tubarão}$_N$ HAS AS A CHARACTERISTIC TO BE {carnívoro}$_{Adj}$ {carnivorous}$_{Adj}$ HAS AS A CHARACTERISTIC TO BE {shark}$_N$
	IS RELATED TO	{marinho}$_{Adj}$ IS RELATED TO {mar}$_N$ {marine}$_{Adj}$ IS RELATED TO {sea}$_N$
	HAS MANNER (*HAS MANNER*)	{balançar}$_V$ HAS MANNER {suavemente}$_{Adv}$ {sway}$_V$ HAS MANNER {smoothly}$_{Adv}$
	IS PART OF (*HAS HOLONYM*) IS INDIVIDUATED PART OF (*HAS HOLO PART*) IS PORTION OF (*HAS HOLO PORTION*) IS MEMBER OF (*HAS HOLO MEMBER*) IS SUBSTANCE/MATERIAL OF (*HAS HOLO MADEOF*) IS SUBLOCATION OF (*HAS HOLO LOCATION*)	{tijolo}$_N$ IS PART OF {parede}$_N$ {tijolo}$_N$ IS PART OF {wall}$_N$ {motor}$_N$ IS INDIVIDUATED PART OF {máquina}$_N$ {motor}$_N$ IS INDIVIDUATED PART OF {machine}$_N$ {gomo}$_N$ IS PORTION OF {citrino}$_N$ {[part of citris fruit]}$_N$ IS PORTION OF {citrus fruit}$_N$ {sócio}$_N$ IS MEMBER OF {clube}$_N$ {member}$_N$ IS MEMBER OF {club}$_N$ {fibra}$_N$ IS SUBSTANCE/MATERIAL OF {tecido}$_N$ {fiber}$_N$ IS SUBSTANCE/MATERIAL OF {tissue}$_N$ {lugar}$_N$ IS SUBLOCATION OF {parque de estacionamento}$_N$ {parking space}$_N$ IS SUBLOCATION OF {parking lot}$_N$
	IS SUBEVENT OF (*IS SUBEVENT OF*)	{inspirar}$_V$ IS SUBEVENT OF {respirar}$_V$ {inspire}$_V$ IS SUBEVENT OF {breathe}$_V$
	IS TELIC SUBEVENT OF	{desconfiado}$_{Adj}$ IS TELIC SUBEVENT OF {desconfiar}$_V$ {suspicious}$_{Adj}$ IS TELIC SUBEVENT OF {[be suspicious]}$_V$
	CHARACTERIZES WITH REGARD TO (*IS VALUE OF*)	{grande}$_{Adj}$ CHARACTERIZES WITH REGARD TO {tamanho}$_N$ {big}$_{Adj}$ CHARACTERIZES WITH REGARD TO {size}$_N$
	CO RELATES WITH (*CO ROLE*)	{telha}$_N$ CO RELATES WITH {ripado}$_N$ {roof tile}$_N$ CO RELATES WITH {roof lattice}$_N$
	CAUSES (*CAUSES*)	{matar}$_V$ CAUSES {morrer}$_V$ {kill}$_V$ CAUSES {die}$_V$
	HAS AS RESULT (*INVOLVED RESULT*)	{construir}$_V$ RESULTS IN/ORIGINATES {construção}$_N$ {build}$_V$ RESULTS IN/ORIGINATES {building}$_N$
	HAS AS LOCATION (*INVOLVED LOCATION*)	{nadar}$_V$ HAS AS LOCATION {líquido}$_N$ {swim}$_V$ HAS AS LOCATION {liquid}$_N$
	HAS AS SOURCE LOCATION (*INVOLVED SOURCE DIRECTION*)	{desencaixotar}$_V$ HAS AS SOURCE LOCATION {caixote}$_N$ {unbox}$_V$ HAS AS SOURCE LOCATION {box}$_N$
	HAS AS GOAL LOCATION (*ROLE TARGET DIRECTION*)	{emigrar}$_V$ HAS AS GOAL LOCATION {estrangeiro}$_N$ {emigrate}$_V$ HAS AS GOAL LOCATION {foreign territory}$_N$
	RESULTS IN/ORIGINATES	{sol}$_N$ RESULTS IN/ORIGINATES {energia solar}$_N$ {sun}$_N$ RESULTS IN/ORIGINATES {solar energy}$_N$
	IS FUNCTION/GOAL OF	{aprender}$_V$ IS FUNCTION/GOAL of {aluno}$_N$ {learn}$_V$ IS FUNCTION/GOAL {pupil}$_N$

Qualia roles	WN.PT Relations	Examples
TELIC	HAS AS FUNCTION/GOAL	{aluno}$_N$ HAS AS FUNCTION/GOAL {aprender}$_V$ {pupil}$_N$ HAS AS FUNCTION/GOAL {learn}$_V$
	HAS TELIC SUBEVENT	{entristecer}$_V$ HAS TELIC SUBEVENT {triste}$_{Adj}$ {sadden}$_V$ HAS TELIC SUBEVENT {sad}$_{Adj}$
	IS THE INSTRUMENT USED FOR (*ROLE RESULT*)	{caneta}$_N$ IS THE INSTRUMENT USED FOR {escrever}$_V$ {pen}$_N$ IS THE INSTRUMENT USED FOR {write}$_V$
	IS THE LOCATION FOR (*ROLE LOCATION*)	{piscina}$_N$ IS THE LOCATION FOR {nadar}$_V$ {pool}$_N$ IS THE LOCATION FOR {swim}$_V$
	IS THE SOURCE LOCATION OF (*ROLE SOURCE DIRECTION*)	{partida}$_N$ IS THE SOURCE LOCATION OF {corrida}$_N$ {start line}$_N$ IS THE SOURCE LOCATION OF {race}$_N$
	IS THE GOAL LOCATION OF (*ROLE TARGET DIRECTION*)	{prisão}$_N$ IS THE GOAL LOCATION OF {encarcerar}$_V$ {prison}$_N$ IS THE GOAL LOCATION OF {imprison, incarcerate}$_V$
	IS TRANSFORMED IN (*CO PATIENT RESULT*)	{girino}$_N$ IS TRANSFORMED IN {rã}$_N$ {tadpole}$_N$ IS TRANSFORMED IN {frog}$_N$
	IS USED TO OBTAIN (*CO INSTRUMENT RESULT*)	{moinho de café}$_N$ IS USED TO OBTAIN {café em pó}$_N$ {coffee grinder}$_N$ IS USED TO OBTAIN {coffee powder}$_N$
	ACTS TO OBTAIN (*CO AGENT RESULT*)	{lenhador}$_N$ ACTS TO OBTAIN {madeira}$_N$ {lumberjack}$_N$ ACTS TO OBTAIN {wood}$_N$
	RELATES AS AGENT WITH THE OBJECT (*CO AGENT PATIENT*)	{lenhador}$_N$ RELATES AS AGENT WITH THE OBJECT {árvore}$_N$ {lumberjack}$_N$ RELATES AS AGENT WITH THE OBJECT {tree}$_N$
	RELATES AS OBJECT WITH THE AGENT (*CO PATIENT AGENT*)	{tijolo}$_N$ RELATES AS OBJECT WITH THE AGENT {pedreiro}$_N$ {brick}$_N$ RELATES AS OBJECT WITH THE AGENT {mason, bricklayer}$_N$
	USES AS INSTRUMENT (*CO AGENT INSTRUMENT*)	{pedreiro}$_N$ USES AS INSTRUMENT {colher de pedreiro}$_N$ {mason, bricklayer}$_N$ USES AS INSTRUMENT {trowel}$_N$
	IS USED AS INSTRUMENT BY (*CO INSTRUMENT AGENT*)	{colher de pedreiro}$_N$ IS USED AS INSTRUMENT BY {pedreiro}$_N$ {trowel}$_N$ IS USED AS INSTRUMENT BY {mason, bricklayer}$_N$
	RELATES AS OBJECT WITH THE INSTRUMENT (*CO PATIENT INSTRUMENT*)	{argamassa}$_N$ RELATES AS OBJECT WITH THE INSTRUMENT {colher de pedreiro}$_N$ {mortar}$_N$ RELATES AS OBJECT WITH THE INSTRUMENT {trowel}$_N$
	RELATES AS INSTRUMENT WITH THE OBJECT (*CO INSTRUMENT PATIENT*)	{colher de pedreiro}$_N$ RELATES AS IJNSTRUMENT WITH THE OBJECT {argamassa}$_N$ {trowel}$_N$ RELATES AS INSTRUMENT WITH THE OBJECT{mortar}$_N$

Table 1: WN.PT relations grouped by qualia role

New relations, in the table above, are marked in green. The relations are here indicated by the labels used in WN.PT[3], although, as presented in chapter 2 (sections 2.1.2 and 2.1.3), the majority of the relations used are the established by EWN model.

Given the quality of the information stated through each relation, and as it is possible to see in Table 1, the two directions of a same relation can refer to different qualia aspects. For instance, meronymy relations are grouped under the Formal and the Constitutive roles depending on the direction focused: holonyms are stated as values of the Constitutive role (HAS AS PART, HAS AS INDIVIDUATED PART, HAS AS PORTION, HAS AS MEMBER, HAS AS SUBSTANCE/MATERIAL, HAS AS SUBLOCATION) since they refer to the parts or elements that constitute a given object, whereas meronyms are stated as values of the Formal role (IS PART OF, IS INDIVIDUATED PART OF, IS PORTION OF, IS MEMBER OF, IS SUBSTANCE/MATERIAL OF, IS SUBLOCATION OF) since they refer to properties of a given object.

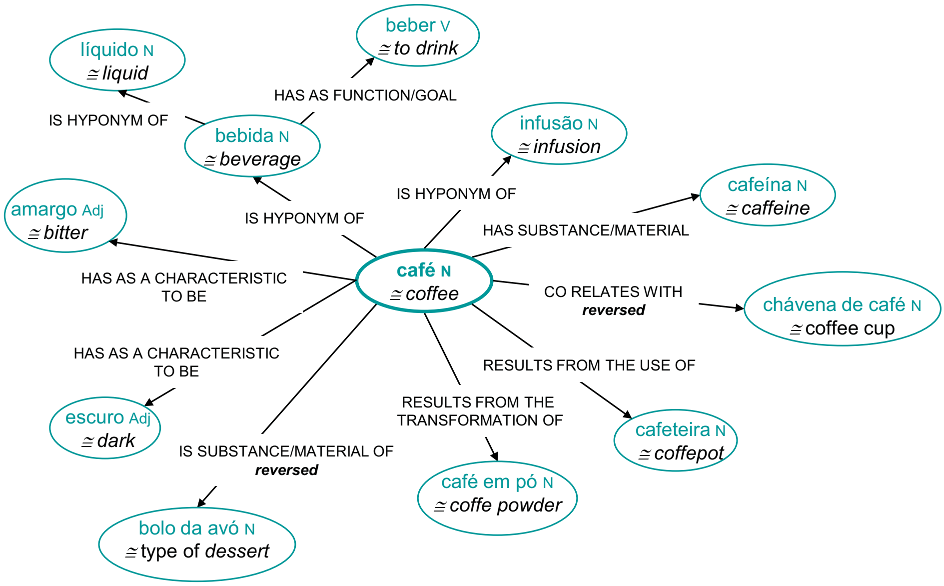
Note, also, that these relations only refer to qualia information (i.e. only constitute values of the qualia roles) if, and only if, they constitute relations not marked by the *reversed* label. That is, only when referring to relations necessary to defining the meaning of a lexical item, these are considered as part of the qualia structure of the lexical items.

For instance, $\{\text{kill}\}_V$ CAUSES $\{\text{die}\}_V$, necessarily. It is an essential defining property of the concept denoted by this synset and, thus, it is expressed in its qualia structure. The reverse is not true: $\{\text{die}\}_V$ is not necessarily CAUSED BY $\{\text{kill}\}_V$ and, thus, this relation is not reflected in the qualia structure associated to the synset $\{\text{die}\}_V$ since it is not an essential defining property of the concept denoted.

The association of the lexical-conceptual relations to the qualia roles they characterize allows the easily recovering the qualia structure of a given item from the set of relations which link it in the net. The network of relations established for the synset $\{\text{café}\}_N$ (≅ coffee), here in (16)a. results in, or expresses, the qualia structure in (16)b.

[3] In WN.PT we adopted informative labels that indicate more clearly the lexical-conceptual relations that connect the nodes in the net, especially considering the use of this resource by non-specialist users.

(16) a. Lexical-conceptual net for {café}$_N$ (≅ coffee)



b. Qualia structure for {café}$_N$ (≅ coffee)

FORMAL = is_hyponym_of(*coffee, beverage*) ∧ is_hyponym_of(*coffee, infusion*) ∧ has_as_a_characterisctic_to_be(*coffee, bitter*) ∧ has_as_a_characteristic_to_be(*coffee, dark*)

CONSTITUVE = has_substance/material(*coffee, caffeine*)

AGENTIVE = results_from_the_use_of(*coffee, coffeepot*) ∧ results_from_the_transformation_of(*coffee, coffee powder*)

The systematic expression of the qualia structure through relations established in EWN model, in this case through the relations established in WN.PT, requires the definition of two new relations: the HAS AS FUNCTION/GOAL(/IS FUNCTION/GOAL OF) and the RESULTS/ORIGINATES FROM(RESULTS IN/ORIGINATES) relations, to express the regular cases of Telic and Agentive properties.

The need for these two relations emerges from the need to systematize the expression of the semantic predicates that may constitute qualia role values. In GL, the values of the Formal, Agentive and Telic roles, for instance *book* = [QUALIA = [FORMAL = book(x), AGENTIVE = write(e_1,y,x), TELIC = read(e_2, y, x), CONSTITUTIVE = has_part(x, y:page), ...]], implicitly state the relations book *is_Formal* book, book *is_Agentive* write, book *is_Telic* read. The Constitutive role values, on the other hand, state explicitly the type of relations through established semantic predicates: book *has_part* page, etc.

Given that the IS SUBTYPE OF and HAS AS PART relations already express the generic cases of Formal and Constitutive properties, respectively, it is only necessary to define the equivalent relations to express the generic case of Agentive and Telic properties. We propose, thus, the relations RESULTS/ORIGINATES FROM and HAS AS FUNCTION/GOAL to fill in this gap.

(17) RESULTS/ORIGINATES FROM/RESULTS IN/ORIGINATES relation

$\{synset\}_1$	RESULTS/ORIGINATES FROM	$\{synset\}_2$ and
$\{synset\}_2$	RESULTS IN/ORIGINATES	$\{synset\}_1$

iff:

i) $\{synset\}_2$ is the origin (natural or artificial) of $\{synset\}_1$ and $\{synset\}_1$ would not exist without $\{synset\}_2$.

ii) $\{synset\}_1$ results in or originates $\{synset\}_2$

(18) HAS AS FUNCTION/GOAL/IS FUNCTION/GOAL OF relation

$\{synset\}_1$	HAS AS FUNCTION/GOAL	$\{synset\}_2$ and
$\{synset\}_2$	IS FUNCTION/GOAL OF	$\{synset\}_1$

iff:

i) $\{synset\}_2$ is the function or goal of $\{synset\}_2$ and

ii) $\{synset\}_1$ has as function or goal $\{synset\}_2$.

The respective counterparts – RESULTS IN/ORIGINATES and IS FUNCTION/GOAL OF –, by their nature, reflect Formal properties.

A first observation of the relations listed in the table above might lead us to consider that some of them may seem redundant with selection relations. Let us consider, for instance, the case of the HAS MANNER, RESULTS/ORIGINATES FROM and HAS AS FUNCTION/GOAL relations.

In fact, the HAS MANNER relation implies necessarily INCORPORATES relation, since adverbial concepts are only necessarily related to a given event if they constitute an element of their semantic content. Note, however, that the reverse is not true: the INCORPORATES relation may not necessarily correspond to a qualia defining relation:

(19) a. $\{whisper\}_V$ INCORPORATES $\{smoothly\}_{Adv}$ and
$\{whisper\}_V$ HAS MANNER $\{smoothly\}_{Adv}$

b. {drink (*drink alcoholic drinks; be alcoholic*)}$_V$ INCORPORATES {alcoholic beverage}$_N$, but
{alcoholic beverage}$_N$ HAS AS FUNCTION/GOAL {drink (*drink alcoholic drinks; be alcoholic*)}$_V$: **false**

The same occurs with RESULTS/ORIGINATES FROM and HAS AS FUNCTION/GOAL relations. Let us consider, for instance, the HAS AS FUNCTION/GOAL relation of the node $\{pupil\}_N$ and the argument structure of the node $\{learn\}_V$.

(20) a. $\{pupil\}_N$ HAS AS FUNCTION/GOAL $\{learn\}_V$

b. $\{learn\}_V$ *SELECTS 1* $\{person\}_N$

SELECTS 2 $\{matter\}_N$ disj: 1:1

SELECTS 2 $\{to\}_{Prep}$ [≅ indicator of theme/subject] disj: 2:1

IS FUNCTION/GOAL OF $\{pupil\}_N$

As it is possible to see, the verb *learn* does not select necessarily for a *pupil* type of argument. As established in (20)b, the first argument of the verb can be realized by $\{person\}_N$ (or by $\{animal\}_N$) and by any of its hyponyms, (being $\{pupil\}_N$ a hyponym of $\{person\}_N$). However, the event denoted by $\{learn\}_V$ is the function or goal of $\{pupil\}_N$, i.e., the entities denoted by $\{pupil\}_N$ have as a distinctive semantic property to have as function or goal $\{learn\}_V$. That is, selection and qualia relations are not redundant since they refer to different levels of meaning representation that are not necessarily mirrored in each other.

Role relations (in (3)) and co-role relations (i.e. relations established between participants of a given event: RELATES AS AGENT WITH THE OBJECT/RELATES AS OBJECT WITH THE AGENT, USES AS INSTRUMENT/IS USED AS INSTRUMENT BY, RELATES AS AGENT/CAUSE WITH THE RESULT/RELATES AS RESULT WITH THE AGENT/CAUSE; RELATES AS INSTRUMENT WITH THE OBJECT/RELATES AS OBJECT WITH THE INSTRUMENT, IS TRANSFORMED IN/RESULTS FROM THE TRANSFORMATION OF, IS OBTAINED THROUGH THE USE OF/IS USED AS INSTRUMENT TO OBTAIN and CO RELATES WITH) also reflect this distinction between levels of representation, especially in the cases where there is no lexicalized specific event that intermediates between the two participants, such as it happens in the examples presented for co-role relations in the table above.

The relations HAS AS FUNCTION/GOAL and RESULTS/ORIGINATES FROM also distinguish stage-level nominals such as *passenger* from individual-level nominals such as *surgeon*. Stage-level nominals can be described as "situationally-defined" nominals, since the concepts they denote result from engaging in a given activity at a given time, i.e. a *passenger* is only a *passenger* while travelling. Individual-level nominals, on the other hand, are "role-defining", in the sense that the concepts they denote describe the function or role of a given agent, regardless of the specific situation he is in, at a given time, i.e. a *surgeon* is still a *surgeon* even when not operating (see Pustejovsky (1995: 229-230)). In both cases the nodes $\{passenger\}_N$ and $\{surgeon\}_N$ can be related to the event denoting nodes $\{travel\}_V$ and $\{operate\}_V$, respectively, through the IS IMPLICATED AS AGENT OF/IMPLICATES AS AGENT (*ROLE AGENT*) relation. However, it is through the relations HAS AS FUNCTION/GOAL and RESULTS/ORIGINATES FROM that it is possible to distinguish the specific semantic properties of these two types of nominals, namely that the first is a stage-level nominal whereas the second is a individual-level nominal:

(21) a. $\{\text{passenger}\}_N$ IS IMPLICATED AS AGENT IN $\{\text{travel}\}_V$, and
$\{\text{passenger}\}_N$ RESULTS/ORIGINATES FROM $\{\text{travel}\}_V$

b. $\{\text{surgeon}\}_N$ IS IMPLICATED AS AGENT IN $\{\text{operate}\}_V$, and
$\{\text{surgeon}\}_N$ HAS AS FUNCTION/GOAL $\{\text{operate}\}_V$

The integration of qualia role in wordnets is, in fact, a simple and low cost process since lexical-conceptual relations in wordnets already reflect intrinsic or prototypical properties that characterize the concepts lexicalized by each synset. This way, the association of the relations to the qualia aspect they refer to allows us to describe the qualia structure of the lexical items in wordnets, without any loss of information and with the advantage of determining the semantic predicates that can be values of the qualia roles in a coherent and consistent way.

7.3 Event structure

The last level of representation to integrate in WN.PT is the event structure. As described in chapter 4 of this work, the event structure is the level of representation that concerns the properties of an event associated to a lexical item. Event structure refers four internal characteristics of the event: its subevents list $<E_1 =..., ..., E_n = ...>$, its Aktionsart type, expressed in GL as a property of each subevent (example: $E_1= e_1$: state), the temporal and order restrictions of its subevents, expressed in GL by the value of the Restrictions attribute, and its Head, attribute that allows the determination of the head subevent.

Event structure is probably the most internal level of representation of the lexical items in the sense that it comprises semantic properties that are not necessarily (or even not at all) related to external elements (being these semantic types in a type lattice or concept denoting synsets in a lexical-conceptual relational net). For these reasons, and contrary to argument and qualia structures, event structure cannot easily be integrated in wordnets as lexical-conceptual relations established between existent nodes, since the properties it defines are not reflected in the nodes in the lexicon.

Given these specificities, we propose the statement of the event structure as additional information at the synset level, through the use of features that mirror the attributes referred above, as well as a new feature that enables the statement of the list of arguments of a given event. As discussed earlier, features convey additional information that is not visible to the system, since it is not expressed through lexical-conceptual relations. Our motivation here is that, nonetheless, the systematic statement of event structure information, besides providing the grounding for argument order description, enriches the descriptive power of the resource making wordnets a rich and structured repository of lexical semantic information that allows the extraction of argument structure and event structure of the lexical items, i.e., the base for a generative lexicon over which devices such as co-composition, selective binding and coercion can operate.

The following table presents the proposed features and respective values that allow the description of event structure.

Features	Values
Event type	State process Transition
Arguments	<1,2,3 …, n>
Subevents	$e_1(2,3)$, …, $e_n(1,2)$
Restrictions	$<$, $\circ$, $<\circ$
Head	$e_{1 \ldots n}$

Table 2: Event structure features

The features here proposed mirror the attributes used in the GL model. The feature **Event type** has as value one of the three possible types of events: state (atomic event, non evaluated regarding any other), process (sequence of identical events (complex or not) that correspond to a process) and transition (event evaluated regarding another event, composed of a process that culminates in a final state, different of the initial one).

Arguments feature has as value the list of arguments selected by the event denoted by the variants of the lexical node, ordered from the less oblique position to the most oblique one. The natural numbers values of this list correspond to the order feature values associated to selection relations, allowing the indexation of the selected nodes to a given position in the list.

The **Subevents** feature allows the listing of the subevents that compose transition denoting events (according to the established typology of events), with information on the arguments of each event. The arguments of each subevent are here referred to by the value of the arguments feature, describing also syntactic realization mapping (see chapter 4, section 4.2).

For instance, a transition type event has as subevents an event argument that corresponds to the process that leads to the final event ($e_1(1,2,3)$) and a second event argument that corresponds to the final state ($e_2(2,3)$). Event type and Subevents features combined express the internal structure of the event (state, process or transition) and the arguments of each subevent.

Restrictions feature allows to express the three possible temporal ordering relations of subevents, established in the GL model: exhaustive ordered part of ($<_{\alpha}$), exhaustive overlap part of ($\circ_{\alpha}$), and exhaustive ordered overlap ($<\circ_{\alpha}$).

And, finally, **Head** feature allows the determination of the head subevent, describing Aktionsart properties (achievements vs. accomplishment type events) as well as events lexically underspecified with regard to event headedness (see chapter 4, section 4.1.1.2).

The slight reformulation of event structure in terms of these five features, established at the synset level, requires also the consequent adjustments on event structure saturation (chapter 4, section 4.2) since the semantic predicates that correspond to the subevents of a complex type denoting event may not be expressed by qualia relations, if they do not correspond to concepts lexicalized in other nodes in the lexicon. For this reason, event structure saturation refers to the head event semantic predicate, expressed not in the qualia roles but in the Subevents feature in the event structure:

(22) EVENT STRUCTURE SATURATION: An event structure is saturated only if all the arguments of the head event semantic predicate, **expressed in the event structure**, are *covered*.

These features allow the expression of the event structure without any loss of information. Note, however, that lexicalized subevents are stated through lexical-conceptual relations at the network level. Thus, verbs that have as subevents conceptually individuated events that are lexicalized, such as {respirar}$_V$ (≅ breathe) in (23)a and {entristecer}$_V$ (≅ sadden) in (23)b, are characterized accordingly through HAS SUBEVENT and HAS TELIC SUBEVENT relations, respectively:

(23) a.

{respirar}v [≅ *breathe*]
Event type: process
Arguments:<1>
Subevents: e1(1)
Restrictions:
Head: e1
Gloss: inspire and expire in alternation to absorve oxigen
Example: Quando o ar é rarefeito, os indivíduos respiram mais rapidamente.
(when the air is rarefied, individuals breathe more quickly)

SELECTS → 1 **{animal}N** [≅ *animal*]
HAS SUBEVENT → **{inspirar}v** [≅ *inspire*]
HAS SUBEVENT → **{expirar}v** [≅ *expire*]
CAUSES → **{respirar}v** [≅ *breathe*; perform gas exchanges]

b.

{entristecer}V [≅ *sadden*]

Event type: transition

Arguments: <1,2>

Subevents: e_1(1,2), e_2(2)

Restrictions: $e_1 < e_2$

Head: e_2

Gloss: make someone sad

Example: A situação entristeceu o rapaz (the situation saddened the boy)

IS HYPONYM OF → **{tornar}V** [≅ *make*]

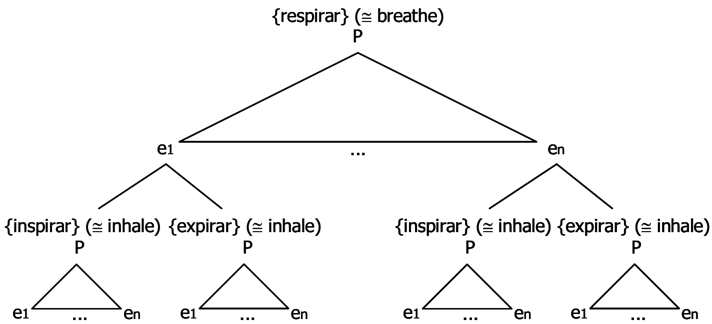
SELECTS → 2 **{person}N** [≅ *pessoa*]

HAS TELIC SUBEVENT → **{triste}Adj** [≅ *sad*]

Given the information stated at the event structure, for instance in the case of the synset {entristecer}$_V$ (≅ sadden), namely that it denotes a transition type event (and thus, a complex event composed of a process type event that leads to a final state event), the telic subevent stated through the HAS TELIC SUBEVENT relation, being a state, corresponds necessarily to the final state event of the transition, whose negation describes the initial state in the sense that it did not exist prior to the event.

In the case of the synset {respirar}$_V$ (≅ breathe), the subevents that compose the process denoted by this verb are described through lexical-conceptual relations since they refer to conceptually individuated lexicalized concepts. However, although being subparts of a process denoting event – a sequence of identical events, i.e., given a time interval I, a process is a sequence of events that is verified in all the intervals of I – the concepts denoted by the nodes {inspirar}$_V$ (≅ inhale) and {expirar}$_V$ (≅ exhale) are subparts of the event that is repeated forming the process (Marrafa 1993:27-28). That is, the process denoted by *respirar* (≅ breathe) is not composed of the sequence of the events *inspirar* (≅ inhale) and *expirar* (≅ exhale), non identical events, but rather of the sequence of identical complex events composed of *inspirar* (≅ inhale) + *expirar* (≅ exhale):

(24) Event structure of {respirar}$_V$ (≅ breathe)



This case further demonstrates the different levels in which event structure properties and lexical-conceptual relations are established, since it is not always the case of direct correspondence between *is part of* relations between event denoting concepts and the internal structure of the events.

The different levels in which argument and qualia structures and event structure are characterized (network level and synset level, respectively) distinguish the different levels of information and allow the coherent representation of the lexical items by differentiating the different kinds of objects used in their characterization.

The addition of event structure information in wordnets involves the fulfillment of the values of the five new features associated to the synset. Given that event type values are pre-established, and that the arguments of the event have to be determined in order to establish selection relations, this task simply requires describing the considered arguments increasing the descriptive power of the resource to extract the syntactic expression of each subevent, when required by context. The following examples illustrate the resultant lexical entries in the network.

(25)

{sair}v [≅ *exit*]
Event type: transition
Arguments: <1,2,3,4>
Subevents: e_1(1,2,3,4), e_2(1,3)
Restrictions: $e_1 < e_2$
Head: e_2
Gloss: move oneself out of
Example: O rapaz saiu do barco. (the boy exited the boat)

IS HYPONYM OF → **{mover-se, deslocar-se}v** [≅ *move oneself*]
INCORPORATES → 2 **{para}Prep** [≅ *to;* direction indicator] conj1:1
INCORPORATES → 2 **{fora}N** [≅ *out*] conj2:1
SELECTS → 3 **{de}Prep** [≅ *from*; source indicator]

(26)

{afastar}$_V$ [≅ *move away*]
Event type: transition
Arguments: <1, 2, 3, 4, 5>
Subevents: e_1(1,2,3,4, 5), e_2(2,4)
Restrictions: $e_1 < e_2$
Head:
Gloss: move away from
Example: O vento afastou o barco do cais. (the wind moved the boat away from the pear)
O barco afastou-se do cais. (the boat moved away from the pear)

IS HYPONYM OF — **{mover, deslocar}$_V$** [≅ *move*]
INCORPORATES — 3 **{para}$_{Prep}$** [≅ *to*; indicator of direction] conj 1:1
INCORPORATES — 3 **{longe}$_N$** [≅ *away*] conj 2:1
SELECTS — 4 **{de}$_{Prep}$** [≅ *from*; source indicator]

7.4 New set of relations and features

As described in the remainder of this chapter, the integration of GL structures in wordnets is achieved through the implementation of a reduced number of relations (five) and features (six). This way, the redefinition of the set of relations used does not result in a large increase of the number of relations used in the model (here, the EWN model). Note, also, that this strategy does not compromise in any way the model since its architecture is completely preserved.

In this section, and to summarize the proposals put forth so far, we present the new set of relations and features we propose to integrate in WN.PT, that in our perspective will enhance greatly, and in a low-cost manner, the descriptive power of this resource.

Table 3, below, presents the relations available in WN.PT, the type of object they relate (word forms, synsets of the same or different POS, specific POS) and the qualia roles they express (A = Agentive, C = Constitutive, F = Formal and T = Telic).

		Relations
Word forms		
		IS SYNONYM OF
		IS ANTONYM OF
Synsets		
Same POS		IS HYPONYM (SUBTYPE) OF (F)/ IS HYPERONYM (SUPERTYPE) OF
		IS NEAR SYNONYM OF
	N	IS PART OF (F)/HAS AS PART (C) IS INDIVIDUATED PART OF (F)/HAS AS INDIVIDUATED PART (C) IS PORTION OF (F)/HAS AS PORTION (C) IS MEMBER OF (F)/HAS AS MEMBER (C) IS SUBSTANCE/MATERIAL OF (F)/HAS AS SUBSTANCE/MATERIAL (C) IS SUBLOCATION OF (F)/ HAS AS SUBLOCATION (C)
		RESULTS FROM THE TRANSFORMATION OF (A)/IS TRANSFORMED IN (T)
		RESULTS FROM THE USE OF (A)/IS USED TO OBTAIN (T)
		RESULTS FROM THE ACTION OF (T)/ACTS TO OBTAIN (T)
		RELATES AS AGENT WITH THE OBJECT(T)/ RELATES AS OBJECT WITH THE AGENT (T)
		USES AS INSTRUMENT/ IS USED AS INSTRUMENT BY (T)
		RELATES AS OBJECT WITH THE INSTRUMENT (T)/ RELATES AS INSTRUMENT WITH THE OBJECT (T)
	V	CAUSES (F)/ IS CAUSED BY (A)
Different POS		IS XPOS NEAR SYNONYM
		IS XPOS NEAR ANTONYM
		SELECTS/ IS SELECTED BY
		SELECTS BY DEFAULT/ IS SELECTED BY DEFAULT
		INCORPORATES/IS INCORPORATED IN
	V/N-N, N-V/N	IMPLICATES AS AGENT/IS IMPLICATED AS AGENT IN
		IMPLICATES AS PATIENT/IS IMPLICATED AS PATIENT IN
		HAS AS RESULT (F)/IS RESULT OF (A)
		HAS AS INSTRUMENT (F)/IS IS THE INSTRUMENT USED FOR (T)
		HAS AS LOCATION (F)/IS THE LOCATION FOR (T)
		HAS AS SOURCE LOCATION (F)/IS THE SOURCE LOCATION OF (T)
		HAS AS GOAL LOCATION (F)/IS THE GOAL LOCATION OF (T)
	Proper N – N, N – Proper N	IS INSTANCE OF/IS INSTANTIATED BY
	V, Adj, N, Adv	HAS SUBEVENT (C)/IS SUBEVENT OF (F)
		HAS TELIC SUBEVENT (C) (T)/ IS TELIC SUBEVENT OF (F)
	V-Adv, Adv-V	HAS MANNER (F)/ IS MANNER OF
	N-Adj, Adj-N	HAS AS A CHARACTERISTIC TO BE (F)/ IS CHARACTERISTIC OF
		CAN BE CHARACTERIZED BY/CHARACTERIZES WITH REGARD TO (F)
		IS RELATED TO (F)

Table 3: Available relations in WN.PT

The next Table lists the set of features and correspondent values available in WN.PT, distinguishing the type of objects they can be applied to: word forms, synsets and relations.

	Features	Values
Word forms		
	Origin	Latin, Spanish, English, French
	Register	Informal, formal, slang, …
	Domain	Scientific, technical, …
Synsets		
	Event type	State, process, transition
	Subevents	e_1 (…, …), … e_n(…, …)
	Restrictions	<, ∘, <∘
	Head	$e_{1…n}$
Relations		
	Reversed	
	Factive	
	Negative	
	Order	1, …, n
	Conjunctive	1…n:1…n
	Disjunctive	1…n:1…n

Table 4: Set of features available in WN.PT

Through this small set of relations and features it is possible to use wordnets as generative lexicons since it allows the modeling and description of all the information required to the operation of generative devices, such as the ones proposed in GL, providing at the same time a rich, accurate and systematic description of the meaning of the lexical items.

7.5 Conclusions

In this chapter we have defined the integration of GL representation levels, namely argument structure, qualia structure and event structure, in a wordnet, demonstrating how wordnets can support a finer-grained lexical description that provides the bases for accounting for several lexical semantic phenomena. This enhancement strategy compensates the emphasis of wordnet model on lexical hierarchy in detriment of the underlying semantics and sustains the use of wordnets for building computational lexica that support generative processes to account for co-composition and meaning in context.

We express argument structure in WN.PT, including default and shadow arguments, by implementing the SELECTS/IS SELECTED BY, INCORPORATES /IS INCORPORATED IN and SELECTS BY DEFAULT /IS SELECTED BY DEFAULT BY relations, that combine with an order feature associated to an argument list for establishing the order of arguments, which, as defined in GL, express constraints on the syntactic mapping. The integration of qualia role in wordnets is achieved by associating lexical-conceptual relations to qualia roles, without any loss of information and with the advantage of determining the semantic predicates that can be values of the qualia roles in a coherent and consistent way. Finally, event structure is expressed through a new set of features (Event type, Arguments, Subevents, Restrictions and Head) that encode the internal properties of the events.

The enhancement of wordnets proposed here reflects the compromise between linguistic knowledge modeling, and thus concerns regarding complexity, and the representation of information useful for processing tasks. This enhancement allows the description of the semantic and syntactic properties of the lexical items, providing the relevant information at the lexical entry level, describing the nature of lexical meaning as well as the specific semantic contribution made by a hyponym in relation to its hyperonym, and assuring the necessary base hierarchy for a default lexical inheritance device, without compromising the global structure of the model.

8. From lexical semantics to syntax

8.0 Introduction

After the presentation of the lexical semantic analysis, that led to the determination of the semantic and syntactic properties considered in the lexical entries of the Portuguese verbs of movement, and of the modeling strategies adopted, this chapter is dedicated to the analysis of some of the constructions in which these verbs occur and the correlation to their lexical semantic properties.

As shown in the introduction of this work, the verbs of movement in Portuguese have different characteristics, namely in what concerns argument structure, Aktionsart properties and the occurrence with expressions denoting the SOURCE and GOAL of the event (also called directed motion constructions, (Gutiérrez 2001:58), the occurrence in middle constructions and the occurrence in causative/non-causative alternations. Our analysis so far has already shown that some of these characteristics are related to the semantic properties that define the concepts denoted by verbs of movement as, for instance, selection properties that mirror different argument structures, internal event structures that account for Aktionsart properties, or semantic properties that account for the occurrence with SOURCE and GOAL denoting expressions. However, some issues related to the type of constituents with which verbs of movement occur, and their occurrence in directed motion constructions, middle and non-causative constructions require further analysis.

This chapter is thus dedicated to the analysis of these issues, aiming at determining if the lexical semantic characterization proposed so far accounts for the different behaviors observed. This analysis is divided in four sections: section 8.1 focuses on the relation between lexical semantics properties and the selection of arguments denoting "region" and "obstacles" occurring in object position (Fong & Fellbaum 2003), and arguments denoting PATH, occurring in oblique positions; section 8.2 is dedicated to the analysis of the expression of directed motion in Portuguese, presenting also a contrastive analysis of the PPs that in Portuguese can convey SOURCE and GOAL locations; section 8.3 deals with the conditions that restrict the occurrence of verbs of movement in middle and non-causative constructions, and section 8.4 is dedicated to the analysis of the clitic *SE* distribution, associated to the passive, middle and non-causative constructions. The final section, section 8.5, summarizes the conclusions of the chapter.

8.1 Obstacle, region and path arguments

As pointed out in the introduction of this dissertation, one of the issues to be considered in the representation of verbs in the lexicon is verbal argument structure and how the determination of the number and type of arguments can be related to the lexical semantic properties of the verbs. Specifically, when it comes to the set of Portuguese verbs of movement, the issue is how lexical semantics properties are related to the selection of arguments denoting obstacle and region, occurring in object position, and arguments denoting path, occurring in oblique positions.

8.1.1 Obstacle and measure

Based on the distinction between obstacle and region arguments, Fong & Fellbaum (2003) propose a three-way typology of verbs of movement:

i) Obstacle-taking verbs, accomplishment denoting verbs that express the traversal of an obstacle in a motion event, but do not specify the manner of motion;

ii) Region-taking verbs, process denoting verbs that express random and unstructured motion; and

iii) Obstacle/Region-taking verbs, verbs that "select for objects that exhibit genuine ambiguity wrt the Obstacle/Region interpretation" (Fong & Fellbaum 2003).

Briefly, the authors characterize obstacle arguments in English by:

i) referring to bounded areas (# *The fighter planes crossed the air*);

ii) not allowing being measured (**John crossed the river for 200 yards*);

iii) entering (periphrastic) passive constructions (*The bridge was crossed*);

iv) entering the middle construction (*This river crosses easily*); and

v) allowing -*ing* nominalizations (*The crossing of the river was difficult*).

Region arguments, on the other hand, are characterized by:

i) referring to unbounded areas (*Mary ambled the countryside*);

i) allowing being measured (*Mary ambled the streets for 3 miles*);

ii) not entering (periphrastic) passive constructions (**The streets were ambled*);

iii) not entering middle constructions (**The new boardwalk ambers easily*); and

iv) not allowing -*ing* nominalization (**?The ambling of the streets was ostentatious*)[1].

Given that the type of argument is determined by the properties of the verb, which is also demonstrated by the fact that some nouns (*street*, for instance) can be interpreted as obstacles or regions depending on the contexts in which they occur, the relevant issue here is the determination of the verbal properties that are reflected in their argument structure.

In the set of Portuguese verbs of movement analyzed, several verbs seem to select an obstacle type argument:

(1) a. atravessar (≅ cross)
O João atravessou a estrada. (John crossed the road.)

b. galgar (≅ pass over)
O cão galgou o muro. (The dog passed over the wall.)

c. orbitar (≅ orbit)
O satélite orbitou o planeta. (The satellite orbited the planet.)

d. circundar (≅ circuit, move around)
O carro circundou a estátua. (The car circuit/moved around the statue.)

e. contornar (≅ circumvent; move near the limits of)
O exército contornou a cidade. (The army circumvented the city.)

f. circum-navegar (≅ circum-navigate)
O navio circum-navegou o continente. (The ship circum-navigated the continent.)

g. percorrer (≅ tour)
Os peregrinos percorreram o santuário. (The pilgrims toured the sanctuary.)

h. escalar (≅ climb)
O João escalou a montanha. (John climbed the mountain.)

However, according to the properties established above, only the verbs in (1)a and b show all the expected properties: incorporate SOURCE & GOAL and PATH (Path in Talmy's typology), respectively, denote accomplishment type events, express the traversal of an obstacle in a motion event, and select arguments that refer to bounded areas, do not allow measurement, and enter passive and middle constructions[2]:

[1] Examples taken from Fong & Fellbaum (2003).
[2] We will not test nominalizations since these are not productive in Portuguese.

(2) a. O João atravessou o rio. (John crossed the river)
O João galgou o muro. (John passed over the wall)

b. #O João atravessou o ar. (#John crossed the air)
#O João galgou o campo. (John passed over the countryside)

c. #O João atravessou 200 metros do rio. (#John crossed 200 meters of the river)
#O João galgou 1 metro do muro. (John passed over/jumped 1 meter of the wall)

d. O rio foi atravessado (pelo João). (The river was crossed (by John))
O muro foi galgado (pelo João). (The wall was passed over (by John))

e. Este rio atravessa-se bem. (This river crosses easily)
Este muro galga-se bem. (This wall passes over easily)

The verbs *orbitar* (≅ orbit), *circundar* (≅ circuit, move around), *contornar* (≅ circumvent; move near the limits of) and *circum-navegar* (≅ circum-navigate) reflect some of these properties, but not all: incorporate PATH and denote accomplishment type events; do not express the traversal of an obstacle in a motion event, but express a motion event directly related to a given reference object (GROUND); select arguments that refer to bounded areas, do not allow measurement and enter passive and middle construction:

(3) a. O satélite orbitou o planeta. (The satellite orbited the planet.)
O navio circum-navegou o continente. (The ship circum-navigated the continent.)
O exército contornou a cidade. (The army moved around/moved near the city limits.)
O carro circundou a estátua. (The car circumbulated the statue.)

b. #O satélite orbitou o espaço. (#The satellite orbited the space.)
#O navio circum-navegou a região. (#The ship circum-navigated the region.)
#O exército contornou o estrangeiro. (#The army circumvented the foreign space.)
#O carro circundou o estrangeiro. (#The car circumbulated the foreign space.)

c. #O satélite orbitou 200 km do planeta. (The satellite orbited 200 km of the planet.)
#O navio circum-navegou 200 km do continente. (#The ship circum-navigated 200 km of the continent.)
#O exército contornou 200 km da cidade. (#The army circumvented 200 km of the city.)
#O carro circundou 20 m da estátua. (#The car circumbulated 20 m of the statue.)

d. O planeta foi orbitado (pelo satélite). (The planet was orbited (by the satellite).)
O continente foi circum-navegado (pelos espanhóis). (The continent was circum-navigated (by the Spanish).)
A cidade foi contornada (pelo exército). (The city was circumvented (by the army).)
A estátua foi circundada (pelo carro). (The statue was circumbulated (by the car).)

e. Este planeta orbita-se bem. (This planet orbits easily.)
Este continente circum-navega-se bem. (This continent circum-navigates easily.)
Esta cidade contorna-se bem. (This city circumvents easily)
Esta estátua circunda-se bem. (This statue circumbulates easily.)

Finally, the verbs *percorrer* (≅ walk around, cover) and *escalar* (≅ climb) seem to be further apart from the properties defined. The verb *percorrer* (≅walk around, cover) denotes an activity type event that lexicalizes MANNER and PATH and selects a location denoting argument. This argument may correspond to bounded or unbounded areas, allows being measured but also enters passive and middle constructions:

(4) a. Os peregrinos percorreram o estrangeiro. (The pilgrims walked around the foreign space)
Os peregrinos percorreram o santuário. (The pilgrims walked around the sanctuary)

c. Os peregrinos percorreram 200 metros do santuário. (The pilgrims walked around 200 meters of the sanctuary)

d. O santuário foi percorrido (pelos peregrinos). (The sanctuary was walked around by the pilgrims)

e. Este santuário percorre-se bem. (This sanctuary walks around easily)

The verb *escalar* (≅ climb), on its turn, denotes an accomplishment type event, selects a GROUND type argument, but incorporates MANNER & GROUND, not PATH, and the argument allows measurement, and enters periphrastic passive and middle constructions:

(5) a. O João escalou a montanha. (John climbed the mountain)

b. #O João escalou a região. (#John climbed the region)

c. O João escalou 200 metros da montanha. (John climbed 200 meters of the mountain.)

d. A montanha foi escalada (pelo João). (The mountain was climbed (by John).)

e. Esta montanha escala-se bem. (This mountain climbs easily)

According to our observation of Portuguese verbs of movement, there are three types of nominal expressions, not corresponding to theme objects, that may occur in object position with these verbs: measure denoting expressions, obstacle/GROUND denoting expressions (the majority of the cases listed above) and region/location denoting expressions (in the case of the verb *percorrer* (≅walk around, cover).

Measure denoting expressions, i.e., expressions referring to the length of the event occur with verbs of all semantic domains and can measure the event in terms of time or in terms of other dimension, directly dependent on the concept denoted by the verb:

(6) a. O João andou 20 metros/20 minutos/durante 20 minutos. (John walked 200 meters/20 minutes/ for 20 minutes)

b. O João estudou 20 páginas/20 minutos/durante 20 minutos. (John studied 200 pages/20 minutes/for 20 minutes)

c. O João dormiu 20 minutos/durante 20 minutos. (John slept 20 minutes/20 minutos/durante 20 minutos)

The measuring of the event sets the limits of the event, having reflexes on the aspectual properties of the sentence. When the verb selects true arguments that already measure the event, or lexicalize semantic components that measure the event, (SOURCE and GOAL, or PATH), typically it does not co-occur with measure denoting expressions, or these apply to the true argument (see (5)c above).

The verbs presented above illustrate this case: verbs that lexicalize SOURCE and GOAL (defining, thus, the initial and final location of the movement) such as *atravessar* (≅ cross), select a defined GROUND argument and do not co-occur with measure denoting NPs; verbs incorporating PATH (defining, thus, the distance of the movement with respect to a reference object) such as *galgar* (≅ pass over) select also a defined GROUND argument and do not co-occur with measure denoting expressions; and verbs incorporating PATH and/or GROUND *orbitar* (≅ orbit), *circundar* (≅ circumbulate), *contornar* (≅ circumvent), *circum-navegar* (≅ circum-navigate) and *escalar* (≅ climb), as discussed in chapter 6, section 6.2.2.1, select GROUND arguments and do not co-occur with measure denoting NPs (*orbitar* (≅ orbit), *circundar* (≅ circumbulate), *contornar* (≅ circumvent), *circum-navegar* (≅ circum-navigate)) or occur with measure denoting NPs that measure the true argument, in the case of the verb *escalar* (≅ climb). Given these properties, it is fair to say that verbs that lexicalize SOURCE & GOAL or PATH select GROUND objects that conform to the properties listed by Fong & Fellbaum (2003) for obstacle-taking verbs. Defined GROUND objects, i.e., denoting concrete and measured entities, correspond to the Obstacle arguments defined by Fong & Fellbaum (2003).

The verb *percorrer* (≅walk around, cover), lexicalizes MANNER and PATH, and selects a location argument that can be measured. Note, however, that in spite of lexicalizing PATH, and thus selecting a GROUND object, the verb *percorrer* (≅ walk around, cover) lexicalizes a strong MANNER component that characterizes the PATH of the event as random.

According to our analysis, monadic change of location verbs (that select one true-argument) (see (7)a) that denote activity type events can co-occur with measure denoting expressions (see (7)b and c.) The event can be measured through the reference to a definite location taken as a space interval, as for instance *a rua toda* (the whole street). These expressions do not necessarily refer to bounded, but measured areas; can be partitioned; can enter middle constructions and can enter some passive constructions:

(7) a. O João andou. (John walked)

b. O João andou 50 metros/ a rua toda.
(John walked 50 meters/the entire street)

c. 50 metros/esta rua toda anda-se bem.
(50 meters/this street walks easily)

e. Andados 50 metros/Andada a rua toda, o João descansou.
(Walked 50 meters/walked the street, John rested)

Although obstacle and measure denoting expressions share many properties, measure expressions do not correspond to true arguments (in (8)), nor can replace obstacle arguments (see (9)):

(8) a. O João andou. (John walked)

b. *O João atravessou. (John crossed)

(9) a. O João atravessou a ponte/*50 metros/*meia ponte.
(John crossed the bridge/*50 meters/half the bridge)

Also, measure expressions can measure the event in terms of time, although only activity denoting change of location verbs that lexicalize MANNER can co-occur with NPs denoting temporal measure:

(10) a. O João nadou 2 horas/o dia todo.
(John swam 2 hours/the whole day)

b. O João avançou *2 horas/?o dia todo/durante 2 horas/durante o dia todo.
(John moved forward 2 hours/?the whole day/for 2 hours/for the whole day)

Given that Portuguese verbs that select GROUND or obstacle arguments do not occur with measure expressions and that measure expressions referring to a definite location (*the whole street*) are not interpreted as regions, the contrast between obstacle and regions is not relevant in Portuguese.

With the exception of the verb *percorrer* (≅walk around, cover), that lexicalizes a strong MANNER component that characterizes the PATH of the event as random, verbs that lexicalize SOURCE & GOAL or PATH select defined GROUND objects, i.e., true arguments denoting concrete and bounded entities syntactically expressed by NPs.

8.1.2 Region and path

In Portuguese, region expressions seem to correspond to PPs introduced by the prepositions *por* or *em* (≅ in/at; indicator of location). However, given that PATH denoting constituents also correspond to PPs introduced by the preposition *por* (≅ through; indicator of path) these expressions can be ambiguous.

Region and PATH denoting constituents are only ambiguous if they do not correspond to true arguments. This way, the distinction between these two types of expressions is defined contextually. In the sentence (11)a, the PP *pela rua* can be interpreted as a PATH or as a region, i.e. an unbounded location. Only in a larger context these expressions are disambiguated (see (11)b and c).

(11) a. O João andou pela rua.
(John walked in the street/through the street)

b. O João andou pela rua$_{[path]}$ até à escola.
(John walked through the street to school)

c. O João andou pela rua$_{[region]}$ durante horas.
(John walked the street for hours)

However, change of location verbs that incorporate restrictions on PATH select a true argument denoting PATH, necessarily introduced by the preposition *por* (≅ through; indicator of path), although some of these verbs can also occur with PPs introduced by the preposition *em* (≅ in/at; indicator of location) (see (12)c and d):

(12) a. {seguir}$_V$ [≅ move for a while through a given path; move along]
O homem seguiu pela/?na estrada velha. (The man moved along through/in the old road).

b. {retroceder}$_V$ [≅ move back through the same path]
O homem retrocedeu pela/*na estrada velha. (The man moved back through/in the old road)

c. {circular, transitar, andar}$_V$ [≅ move usually through a given space/way; circulate].
O homem circulou/transitou/andou pela/?na estrada velha. (The man circulated through/in the old road)

d. {circular}$_V$ [≅ move cyclically, through a given path; circulate]
O sangue circulou pelas/nas artérias. (The blood circulated through/in the arteries)

This different behavior can be related to the MANNER component incorporated in the meaning of the verbs *circular*, *transitar, andar* and *circular*, in (12)c and d, respectively, that conveys a strong iterativity aspect of the movement event (*usually*, in the first case; *cyclically*, in the second), licensing the co-occurrence with location denoting PPs. That is, the movement event happens repeatedly in a given space, and, if no path is specified, these constituents are necessarily interpreted as the path of the movement. Note also that all these verbs may co-occur simultaneously with PATH and location denoting expressions:

(13) a. O homem seguiu [pela estrada velha]$_{PATH}$ [na zona militar]$_{LOCATION}$.
(The man moved along [through the old road] [in the military zone]$_{LOCATION}$).

b. O homem retrocedeu [pela estrada velha] [na zona militar].
(The man moved back [through the road] [in the military zone])

c. O homem circulou/transitou/andou [pela estrada velha]$_{PATH}$ [na zona militar]$_{LOCATION}$.
(The man circulated [through the old road] [in the military zone]$_{LOCATION}$)

d. O sangue circulou [pelas artérias] [na zona abdominal].
(The blood circulated [through the arteries] [in the abdominal zone]$_{LOCATION}$)

These data show that PPs introduced by the preposition *por* (≅ through) refer to PATH, whereas PPs introduced by the preposition *em* (≅ in/at) refer to locations.

On the other hand, Portuguese verbs that can belong to the class of region-taking verbs considered by Fong & Fellbaum (2003), although not as productive in Portuguese as they are in English, correspond to process denoting verbs that express undirected movement through a random and unstructured path and can co-occur with expressions introduced by *por* (≅ through; indicator of path) or *em* (≅ in/at; indicator of location).

Verbs such as {deambular, vaguear}$_V$ (≅ wander) are some of the few examples that occur in Portuguese:

(14) a. O homem deambulou/vagueou pela/na cidade. (The man wandered through/in the city)

In this case, both expressions are interpreted as locations due to the concept denoted by the verb, of undirected movement through a random and unstructured path, which can also explain why these verbs do not easily co-occur with SOURCE and GOAL denoting constituents:

(15) a. ?#O homem deambulou/vagueou pela cidade [da estação]$_{SOURCE}$ [para o aeroporto]$_{GOAL}$.
(The man wandered through the city [from the station]$_{SOURCE}$ [to the airport]$_{GOAL}$)

However, the small number of verbs in this set seems to be insufficient to undoubtedly corroborate this hypothesis.

The co-occurrence of PATH and location expressions with verbs of change of location is necessarily related to the lexical semantic properties of verbs. The lexical semantic analysis of Portuguese verbs of movement presented so far, and as demonstrated throughout the remainder of this work, can thus be decisive for the explanation of different selection and subcategorization properties of lexical-conceptually related verbs, although not always lexical semantic properties result in regular patterns of behavior.

8.2 The directed motion structures

The possibility of occurring in directed motion structures, i.e. with PPs that express the SOURCE and GOAL of the movement, is frequent within the class of verbs of movement. The diverse behavior and co-occurrence restrictions among the verbs of movement class constitute an interesting issue to be addressed, since the diversity of contexts in which verbs occur and the meaning variation that occurrence contexts allow are expected to be related to lexical semantic properties. The treatment of these structures – described as syntactic structures in which the verb occurs with PPs referring the source location (*de* (≅from)+NP) and the goal location (*para/até a* (≅to)+NP) – is usually related to major differences in the modeling and representation of the semantic content of verbs of movement and in the number of lexical entries considered.

Traditionally, the issue is focused on the subclass of manner of motion verbs, where these implications are more visible:

(16) a. O João correu/corre. (John ran/runs): process

b. O João correu [do café]$_{\text{SOURCE}}$ [até a/para a escola]$_{\text{GOAL}}$
(John ran from [the coffee shop]$_{\text{SOURCE}}$ [to school]$_{\text{GOAL}}$): accomplishment

Base on data such as the presented above, the discussion is centered on determining if the verb *correr* (≅ run) – a manner of motion verb – can co-occur with SOURCE and GOAL denoting expressions or if, on the contrary, the two structures correspond to two different verbs with distinct lexical entries and argument structures, mirroring also the aspectual properties that the two sentences above illustrate.

However, there seems to be more issues to be addressed with respect to the occurrence of Portuguese verbs of movement with directional expressions, in particular the definition and treatment of these directed motion structures and the modeling of the information stated in the lexical entries of the verbs.

8.2.1 Approaches to directed motion constructions

The directed motion construction has been object of several studies that can be grouped under a lexical approach (Jackendoff 1983, Van Valin & Lapolla 1997, Rappaport Hovac & Levin 1998), a constructional approach (Goldberg 1995, Goldberg & Jackendoff 2004) and compositional approach (Pustejovsky 1991, Hoekstra 1992).

In a lexical perspective (Jackendoff (1983, 1990), Rappaport Hovac & Levin (1998), Van Valin & Lapolla (1997), etc.), it is considered that the directed motion construction reflects meaning differences that are stated in the lexical entry of the verbs, since "the verb contains all the information about the clausal structures it will head" (Gutierrez 2001: 71). According to this approach, there are two distinct lexical entries for the verb *correr* (≅ run): one corresponding to the one-place verb that denotes a process type event and expresses manner of motion, illustrated in (16)a, and a second one corresponding to a three-place verb that denotes an accomplishment type event and expresses directed movement, in (16)b.

Yet, this approach does not consider the meaning of directed movement that results from the occurrence of PPs such as *de* (from) … *até a* (to) with process denoting verbs that do not express movement events:

(17) a. A Ana leu ininterruptamente de Lisboa até Paris.
(Ana read uninterruptedly from Lisbon to Paris)

b. A Ana fumou do café até à escola.
(Ana smoked from the coffee shop to school)

In a "constructional approach" (Goldberg 1995), it is argued the existence of a "caused-motion" construction, since the semantics of the construction is not derivable from the inherent meaning of the verb. According to Goldberg & Jackendoff (2004), some contexts demonstrate that the presence of some arguments cannot be attributed to the selection properties of the verb, as exemplified in (18), motivating the existence of semi-rigid constructions in the Lexicon. In this approach, thus, constructions must be considered in the lexicon.

(18) a. The professor talked us into a stupor.
b. *The professor talked us.

In compositional approaches, both the importance of the meaning of a given verb and of its interaction with the elements of the syntactic construction in which it occurs are recognized. Gutiérrez (2001), for instance, argues that the meaning of directed motion sentences results from "the integration of the meaning of the verbs with the meaning of the constructions in which they appear, but at the same time the meaning of the verb is determined and can be changed to a great extent by the particular syntactic configurations in which it participates more frequently, giving rise to new lexical entries for that verb" (Gutiérrez 2001:71). Pustejovsky (1991), Hoekstra (1992) or Zubizarreta & Oh (2004) consider, on the other hand, that the meaning of a sentence is compositional, i.e. results from the combination of the meanings of the elements of the sentence, being the possibility of a given verb entering certain constructions determined lexically.

8.2.2 Directed motion structures in Portuguese

It is not clear that in Portuguese there is a rigid or semi-rigid directed motion structure. The determination of a rigid structure does not account for Portuguese verbs of movement behavior for several reasons:

a) verbs of movement semantically legitimate SOURCE and GOAL locations complements, in the same way that they legitimate the occurrence with PATH constituents, for instance, not traditionally considered expressed in the constituents that form directed motion structures (*de* (≅ from) and *até/para* (≅ to)):

(19) a. O capitão avançou [para o convés]$_{\text{GOAL}}$.
(The captain move forward [to the deck]$_{\text{GOAL}}$)

b. O capitão correu [da proa]$_{\text{SOURCE}}$ [para o convés]$_{\text{GOAL}}$
(The captain ran [from the stem]$_{\text{SOURCE}}$ [to the deck]$_{\text{GOAL}}$)

c. O capitão desceu [do convés]$_{\text{SOURCE}}$ [pelas escadas]$_{\text{PATH}}$.
(The captain descended [from the deck]$_{\text{SOURCE}}$ [through the stairs]$_{\text{PATH}}$)

d. O navio foi [para norte]$_{\text{DIRECTION}}$.
(The ship went [to north]$_{\text{DIRECTION}}$)

b) The expressions denoting SOURCE and GOAL are part of the argument structure of some verbs of movement, and thus are not part of a rigid directed motion construction:

(20) a. O vento afastou o navio da margem.
(The wind moved away the ship from the shore)

a'. ??O vento afastou o navio.
(The wind moved away the ship.)

b. O capitão foi para o convés.
(The captain went to the deck.)

b'. *O capitão foi.
(The captain went.)

c. O capitão esgueirou-se do navio.
(The captain sneaked out of the ship)

c'. ?*O capitão esgueirou-se.
(The captain sneaked out.)

c) The elements of the directed motion construction *de* (≅ from) and *até/para* (≅ to) are not required to co-occur simultaneously, that is, these expressions may occur alone. Also, several co-occurrence restrictions with these expressions take place with some verbs of movement:

(21) a. O navio avançou [para o cais]$_{\text{GOAL}}$.
(The ship move forward [to the pear]$_{\text{GOAL}}$).

b. O navio avançou [da baía]$_{\text{SOURCE}}$ [para o cais]$_{\text{GOAL}}$ [pelo canal]$_{\text{PATH}}$.
(The ship moved forward [from the bay]$_{\text{SOURCE}}$ [to the pear]$_{\text{GOAL}}$ [through the canal]$_{\text{PATH}}$)

c. O capitão aproximou-se [da ponte]$_{\text{GOAL}}$
(The captain moved closer [to the bridge]$_{\text{GOAL}}$)

c'. *O capitão aproximou-se [do convés]$_{\text{SOURCE}}$ [para a ponte]$_{\text{GOAL}}$.
(The captain moved closer [from the deck]$_{\text{SOURCE}}$ [to the bridge]$_{\text{GOAL}}$)

d. *O capitão subiu [para a ponte]$_{\text{GOAL}}$ [para baixo]$_{\text{DIRECTION}}$
(The captain went up/ascended [to the bridge]$_{\text{GOAL}}$ [downwards]$_{\text{DIRECTION}}$)

According to our analysis, and as already explained in section 8.1, the information present in the lexical entries of the items that denote SOURCE, GOAL, PATH and DIRECTION concepts license their occurrence with verbs of movement. On their turn, the properties in the lexical entries of verbs of change of location, license and/or restrict their co-occurrence with SOURCE, GOAL, PATH and DIRECTION denoting constituents, according to the semantic elements incorporated in the verbs and to their subcategorization properties.

The decompositional analysis of the meaning of verbs of movement allow us to predict which verbs may occur with which constituents, without devising lexical entries for a rigid directed motion construction, which, as we have demonstrated, is not motivated.

Some verbs denote a given direction and in that case, naturally, can only occur with constituents that denote a compatible direction. For instance, verbs that incorporate DIRECTION, such as *afastar-se* (≅ move away) or *descer* (≅ descend, move down) easily co-occur with PPs denoting SOURCE, GOAL or PATH ((22)a), but only co-occur with PPs denoting DIRECTION if these are compatible with the DIRECTION denoted by the verb (see (22)b and c):

(22) a. O capitão avançou [da popa]$_{\text{SOURCE}}$ [para o convés]$_{\text{GOAL}}$ [pelo passadiço]$_{\text{PATH}}$
(The captain moved forward [from the stern]$_{\text{SOURCE}}$ [to the deck]$_{\text{GOAL}}$ [through the aisle]$_{\text{PATH}}$)

b. *O capitão avançou [para trás]$_{\text{DIRECTION}}$.
(The captain moved forward [backwards]$_{\text{DIRECTION}}$)

c. O capitão avançou [para a esquerda]$_{\text{DIRECTION}}$.
(The captain moved forward [to the left]$_{\text{DIRECTION}}$)

Besides accounting for the semantic licensing of the co-occurrence with these constituents, argument structure also accounts for the co-occurrence restrictions directly related to subcategorization specificities, as in the case of the verb *aproximar-se* (≅ move closer to), illustrated in (21)c and c' above. The verb *aproximar-se* (≅ move closer to) selects a true argument denoting goal location, but introduced by the argument-marking preposition *de*, homographous of the preposition denoting indicator of source location. For this reason, the PP is necessarily interpreted as goal indicator, and cannot occur with GOAL denoting PPs introduced by *para*, given that the GOAL of the event is already established by the true argument of the verb.

For the reasons presented so far, we follow a compositional approach since we consider that the meaning of a sentence in which a given verb of movement occurs is function of the relation between the meaning of its elements, being the co-occurrence restrictions of the verbs with certain type of expressions determined at the lexical level. This analysis imposes the independent treatment of the prepositions that introduce the constituents that form directioned motion structures, and not the establishment of a lexical entry for a given construction.

8.2.3 *de* (≅ from) *... para* (≅ to) vs. *de/desde* (≅from/since) ... *até a/a* (≅ until)

One other observation to be made regarding directed motion structures in Portuguese is related to the use of PPs introduced by the preposition *a*, considered in research on English, and the occurrence restrictions of this PPs with manner of motion verbs in Romance languages, as, for instance, Spanish (Zubizarreta & Oh 2004:25) and Portuguese:

(23) a. The athletes swam to the boat.
a'. *Los atletas nadaron al barco.
a''. *Os atletas nadaram ao barco.

b. The bottle floated to the beach.
b'. *La botella flotó a la playa.
b''. *A garrafa flutuou à praia.

In fact, it seems that these manner of motion verbs do not occur with this preposition in Portuguese, contrary to what happens with other verbs of movement, such as *ir* (≅ go), *vir* (≅ come) or *voltar* (≅ return).

(24) a. O João foi à cozinha. (John went to the kitchen)

b. Os atletas voltaram ao barco. (The athletes returned to the boat)

c. O atleta veio à praia. (The athlete came to the beach)

Based in this observation, Zubizarreta & Oh (2004) propose that manner of motion verbs are not true verbs of movement, that is, that these verbs do not denote change of location, being this meaning obtained compositionally, as it happens with verbs that do not denote movement events (see (17) above).

Note, however, that in Portuguese there are verbs of movement, other than manner of motion verbs, that cannot occur with PPs introduced by *a*, including the top node verb *mover, deslocar* (≅ move, change location), and that, nonetheless, denote change location:

(25) a.*O vento moveu/deslocou o barco à margem.
(The wind moved the boat to the shore)

b. *O vento avançou[+DIR.] o barco ao cais.
(The wind moved forward the boat to the pear)

In our opinion, the Portuguese preposition *a* is not the natural correspondent of the English preposition *to*, indicator of goal location.

In Portuguese, verbs of movement, including the subclass of manner of motion verbs, may generally occur with different structures indicating source and goal locations:

i) *de/desde* (≅ from/since)+NP *até a*(≅ to/until)+NP, and

ii) *de*+NP(≅ from)+NP *para*+NP (≅ to)+NP

In the majority of cases, the expressions of these structures are interchangeable without any major change in the meaning of the sentence:

(26) a. O atleta nadou de/desde o barco até à praia.
(The athlete swam from/since the boat to/until the beach)

b. O atleta nadou do barco para a praia.
(The athlete swam from the boat to the beach)

However, only the PPs introduced by *de/desde* (≅ from/since) and *até a*(≅ to/until) can occur with verbs from other semantic classes that do not denote change of location, such as *fumar* (≅ smoke) or *ler* (≅ read), as the sentences in (27) illustrate. Note also that when occurring with process denoting verbs, these expressions bound the denoted event, imposing a limit to the denoted process and changing the sentence aspectual value to that of an accomplishment.

(27) a. A Ana leu ininterruptamente de Lisboa até Paris.
(Ana read uninterruptedly from/since Lisbon to/until Paris.)

a'. ?*A Ana leu ininterruptamente de Lisboa para Paris.
(Ana read uninterruptedly from Lisbon to Paris)

b. A Ana fumou do café até à escola.
(Ana smoked from/since the coffee shop to/until the school)

b'. ?*A Ana fumou do café para a escola.
(Ana smoked from the coffee shop to the school)

These data show that the construction *de/desde* (≅ from/since) … *até a* (≅ to/until) delimits the event denoted by the verb, and that these limits may be expressed by temporal or spatial expressions.

The restrictions related to the occurrence with these expressions, especially in what concerns the type of NP (referring to time or space) selected by the preposition, are related to the semantic properties of the verbs with which they co-occur, as well as with the verbal selection properties. Consider, for instance, the verb *ir* (≅ go), that denotes an accomplishment type event and selects a true argument denoting GOAL, in (28), as opposed to the verb *avançar* (≅

move forward), that denotes a process type event and that does not select for SOURCE or GOAL denoting arguments, in (29).

(28) a. O capitão foi para/até ao convés.
(The captain went to/until the deck)

b. O capitão foi desde a/da popa para/até ao convés.
(The captain went from/since the stern to/until the deck)

c. *O capitão foi até às 6 horas.
(The captain went until 6 o'clock)

d. O capitão foi para o convés das 3 às 6 horas.
(The captain went to the deck from 3 until 6 o'clock)

(29) a. O capitão avançou para/até ao convés.
(The captain moved forward to/until the deck)

b. O capitão avançou da popa para/até ao convés.
(The captain moved forward from/since the stern to/until the deck)

c. O capitão avançou até às 18 horas.
(The captain move forward until 6 o'clock)

d. O capitão avançou das 15 às 18 horas.
(The captain move forward to the deck from 3 until 6 o'clock)

The example in (28)c shows that a temporal argument cannot replace a spatial argument. On the other hand, (28)d reading is that the final state of the accomplishment event is that the captain stood in the deck for 3 hours, and not that the change of location event took 3 hours to be accomplished. The sentences in (29) show that the occurrence of the verb *avançar* (≅ move forward) with phrases denoting whether goal, whether a time limit ((29)a and c, respectively) results in an accomplishment reading, given the delimitation of the process denoted by the verb.

The observation of the data confirm that the occurrence of verbal predicates with the expressions *de/desde* (≅ from/since) … *até a* (≅ to/until) is licensed by the Aktionsart properties of the verbs, regardless of the semantic domain they belong to. Simultaneously, it becomes clear that the occurrence with PPs denoting SOURCE and GOAL is licensed by the semantic properties of the verbs, that is, by the fact of these verbs denoting change of location, in the same way that that the occurrence with PPs denoting PATH is licensed by change of location verbs, for instance, while other verbs rarely allow this co-occurrence and when they do it does express the PATH of the event:

(30) a. A Ana leu pelo corredor. (≅ A Ana esteve em vários pontos do corredor & A Ana leu)
(Ana read in the corridor.(≅ Ana was in several points of the corridor & Ana read))

In (30), the PP is interpreted as denoting locations and not as denoting the PATH of the reading event.

The distinction between these two structures, however, does not explain why some Portuguese verbs of movement can occur with the preposition *a*, while others cannot.

The distribution of the preposition *a* denoting a final location (instead of the expression *até a* (≅ until)) considering the set of verbs of movement analyzed seems to be related to restrictions of several order. Within the set of Portuguese verbs denoting change of location, it is possible to observe that:

i) verbs selecting PPs expressing PATH, typically, do not occur with PPs introduced by *a*, regardless of the semantic elements they incorporate:

(31) a. $\{seguir\}_V$ (≅ move along; move +PATH)
*O João seguiu pela estrada à escola. (John moved along the road to the school)

ii) one-place verbs denoting manner of motion, typically, do not occur with PPs introduced by the preposition *a*:

(32) a. $\{andar, caminhar\}_V$ (≅ walk; move +MANNER)
*O João andou/caminhou à escola. (John walked to the school)

b. $\{coxear, manquejar, mancar\}_V$ (≅ limp; move +MANNER)
*O João coxeou/manquejou/mancou à escola. (John limped to the school)

c. $\{patinar\}_V$ (≅ skate; move +MANNER)
*O João patinou à escola. (John skated to the school)

On the other hand, PPs introduced by *a* denoting goal location occur with verbs that do not select NPs in object position incorporating DIRECTION of movement, in (33),

(33) a. $\{descer\}_V$ (≅ move down, descend; move +DIRECTION)
O João desceu à cave. (John moved down to the basement)

b. $\{subir\}_V$ (≅ move up, ascend; move +DIRECTION)
O João subiu ao sótão. (John moved up to the attic)

and can replace the goal denoting argument selected by accomplishment and achievement denoting verbs incorporating DIRECTION and GOAL of movement, in (34):

(34) a. {ir, deslocar-se}$_V$ (≅ go; move + DIRECTION+GOAL)
O João foi/deslocou-se à cave. (John went to the room door)

b. {vir}$_V$ (≅ come; move +DIRECTION+GOAL)
O João veio à cave. (John came to the basement)

c. {trepar, amarinhar, marinhar}$_V$ (≅ climb; move up; move +DIRECTION+MANNER)
O João trepou/amarinhou/marinhou ao sótão. (John climb to the attic)

However, not all the verbs incorporating DIRECTION can occur in this construction (in (35)a), on the one hand, and some verbs denoting manner of motion occur in this context (in (35)b to e), on the other:

(35) a. {avançar}$_V$ (≅ move forward; move +DIRECTION)
*O João avançou à cave. (John moved forward to the basement)

b. {correr}$_V$ (≅ run; move +MANNER)
O João correu ao hospital. (John ran to the hospital)

c. {arrastar}$_V$ (≅ drag; move +MANNER)
O João arrastou a Maria à cave. (John drag Mary to the basement)

d. {rastejar}$_V$ (≅ crawl; move +MANNER)
?O João rastejou à cave. (John crawled to the basement)

e. {voar}$_V$ (≅ fly; move +MANNER + GROUND)
O João voou à Venezuela. (John flew to Venezuela)

The data show that the distribution of goal denoting PPs introduced by *a* is conditioned by the type of movement event denoted by the verb (manner of motion vs. directed motion), but also by Aktionsart properties and other use factors (some manner of motion denoting verbs occur with this PPs, other do not). Also, the occurrence of this PP with source denoting PPs in simultaneous seems to improve the acceptability of some contexts:

(36) a. {mover-se}$_V$ (≅ move oneself; move +FIG)
*O João moveu-se ao quarto. (John moved to the bedroom)
O João moveu-se da cozinha ao quarto. (John moved from the kitchen to the bedroom)

b. {andar, caminhar}$_V$ (≅ walk; move +MANNER)
*O João andou/caminhou à escola. (John walk to the school)
O João andou/caminhou de casa à escola. (John walk from home to the school)

c. {patinar}$_V$ (≅ skate; move +MANNER)
*O João patinou à escola. (John skated to the school)

O João patinou de casa à escola. (John skated from home to the school)

This seems to indicate that, in these contexts, *a* is the shortened version of the prepositional expression *até a* (≅ until) and, thus, more easily recovered when the source location is also indicated.

However, apart from this specific case, there seems to be also some meaning differences that oppose the expression of goal locations through PPs introduced by *a* to the expression of goal locations through PPs introduced by *para* (≅ to) or even by *até a* (≅ to/until). PPs introduced by *a* induce a punctual aspect interpretation of the final state of the event:

(37) a. O João foi ao Porto. (≅ John went to Oporto but is no longer there)

b. O João foi para/até ao Porto. (≅ John went to/until Oporto and is still there)

This hypothesis predicts that the preposition *a* distribution is conditioned by the Aktionsart properties of these verbs: expressions denoting goal location introduced by the preposition *a* can occur with accomplishment or achievement denoting verbs of change of location, explaining why verbs such as *avançar* (≅ move forward; move +DIRECTION) and *mover-se* (≅ move oneself; move +FIGURE), activities denoting verbs, cannot co-occur with these expressions.

Also, the exceptions listed in (35)b to e, of manner of motion verbs that occur with goal expressions introduced by the preposition *a*, seem to be licensed in specific contexts where the verb denotes go/take+behavior/manner/means, i.e, accomplishment type events:

(38) a. O João correu ao hospital. (≅ O João foi com urgência ao hospital)
(John run to the hospital (≅ John went with urgency to the hospital))

b. O João arrastou a Maria à cave. (≅ O João levou a Maria à cave contra a sua vontade)
(John dragged Mary to the basement (≅ John took Mary to the basement against her will))

b'. *O João arrastou a caixa à cave. (John dragged the box to the basement)

c. ?O João rastejou à cave. (≅ O João foi à cave com muitas dificuldade)
(John crawled to the basement (≅ John went to the basement with extreme difficulty))

d. O João voou à Venezuela. (≅ O João foi à Venezuela de avião)
(John flew to Venezuela (≅ John went to Venezuela by plain)

d'. *O pássaro voou à Venezuela. (The bird flew to Venezuela)

The distribution of the preposition *a*, indicator of goal location, with verbs of movement in Portuguese requires a thorough analysis that is out of the scope of this dissertation. However, the data presented here show that the proposal of Zubizarreta & Oh (2004) that manner of motion verbs are not true verbs of movement, based on the co-occurrence with this particular preposition, does not account for the Portuguese case. As mentioned earlier in this section, Portuguese verbs of movement denoting manner of motion events can in fact occur with PPs denoting source and goal locations introduced by *de* (≅ from; indicator of source location) and *para* (≅ to; indicator of goal location) or by *de, desde* (≅ from; indicator of beginning) or *até a* (≅ until; indicator of end), and are thus seen as true verbs of movement (as also defended by Croft (2000: 79-82)).

8.3 Middle and non-causative constructions

Middle construction and causative/non-causative alternations constitute constructions in which verbs of movement typically occur.

The middle construction can be characterized by its generic use (Ruwet 1972) but also by not occurring with adverbs that specify speech point and not allowing conditional interpretations in absolute constructions – that is, middle construction describes generic or stative properties –; and not occurring with overtly expressed subjects (although necessarily understood (see Fong *et al.* (2000) and Kageyama (2002), for instance)).

This construction, according to several authors (Jaeggli 1986, Tenny 1987, Hoekstra & Roberts 1993, Fagan 1992), is only possible if the predicates internal object is "affected", i.e. if the object is somewhat affected by the event denoted by the verb, or, within thematic role theory, receives a Patient thematic role). However, this assumption is not consensual (Fong *et. al* 2000) and it also does not apply to several Romance languages (see Ruwet 1972, Zribi-Hertz 1987, Cinque 1988, Cornips & Hulk 1999, Fong *et al.* (2000), Hulk & Cornips 2000), Portuguese included.

(39) a. Spy novels read easily. (example from Fong *et. al* (2000))

b. Romances de espiões lêem-se facilmente.
novels of spies read-SE easily

Portuguese verbs of movement that select obstacle, ground and measure arguments – not affected objects –, for instance, allow for middle constructions:

(40) a. Pontes metálicas atravessam-se facilmente. (Metal bridges cross easily)

b. Estas rampas escalam-se facilmente. (These ramps climb easily)

c. Estes planetas orbitam-se facilmente. (These planets orbit easily)

d. 100 metros correm-se facilmente. (100 meters run easily)

In general, Portuguese verbs of movement that take arguments in object position enter middle construction:

(41) a. Crianças pequenas embalam-se facilmente.
(Small children rock easily)

b. Livros altos inclinam-se facilmente nas prateleiras.
(Tall books lean on shelves easily)

c. Crianças pequenas deitam-se facilmente.
(Small children lay easily)

The causative/non-causative alternation is also quite productive in Portuguese. Non-causative constructions include, but are not restricted to, inchoative constructions, typically allowed by verbs of change of state whose object is affected by the event denoted by the verb. In the non-causative construction, the object occurs in the syntactic subject position. Non-causative constructions denote an event in which the mention of an agent or cause is taken as irrelevant, although possible (cf. Levin (1993: 30). In Portuguese, these are expressed through PPs introduced by *com* (≅ with; indicator of cause).

Given the different syntactic realizations of causative vs. non-causative constructions and the consequences these alternative constructions have in the determination of the number arguments of verbs and of the information stated at the lexical entries, Pustejovsky (1995) relates the internal properties of complex events such as transitions, namely the headedness of events, to causative and inchoative alternations. According to this author, causative/inchoative alternations are possible with verbs that denote transitions whose head event is not lexically defined. This way, if the head event is the final state, the argument structure projected is the one associated to this subevent; if the head event is the preparatory process, the argument structure projected is the one associated to that subevent:

(42) a. *sink* :

$$\begin{bmatrix} \text{EVENTSTR} = \begin{bmatrix} \text{E1} = \text{e1} : \text{process} \\ \text{E2} = \text{e2} : \text{state} \\ \text{RESTR} = < \alpha \end{bmatrix} \\ \text{QUALIA} = \begin{bmatrix} \text{AGENTIVE} = \text{sink\_act}(\text{e1}, \text{x}, \text{y}) \\ \text{FORMAL} = \text{sink\_state}(\text{e2}, \text{y}) \end{bmatrix} \end{bmatrix}$$

b. The navy sunk the ship. (head event = process: sink(e_1, x, y))

c. The ship sunk. (head event = final state: sink_state(e_2, y))

Event headedness accounts, thus, for the causative and inchoative alternations.

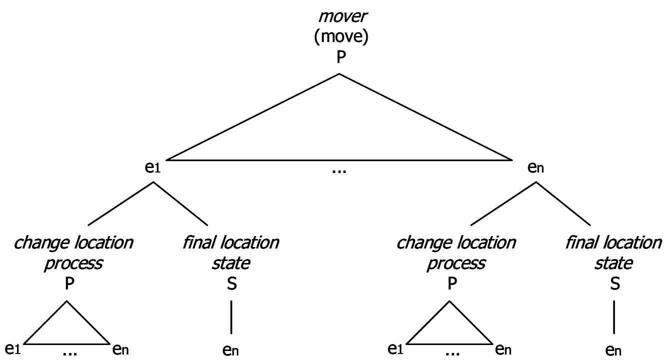
Verbs of movement are not usually considered with regard to causative/inchoative alternations, although these verbs do occur in these constructions:

(43) a. O vento moveu/deslocou o balde (The wind moved the bucket)

b. O balde moveu-se/deslocou-se (com o vento). (The bucket moved (with the wind)

However, and as previously discussed (see chapter 6, section 6.1), some verbs of movement denote processes (as the Aktionsart tests in (45) show), although these correspond to complex processes composed of the repetition of transition type events:

(44) Event structure for the verb *mover* (≅ move)



(45) a. O homem moveu/deslocou o objecto durante 10 minutos.
(The man moved the object for 10 minutes)

b. O homem está (agora) a mover/deslocar o objecto → o homem (já) moveu/deslocou o objecto.
(The man is (now) moving the object→ the man has (already) moved the object)

This structure, as discussed in chapter 4, sections 4.1.1.2 and 4.1.1.3, is not directly reflected in the event structure level of representation considered in this work since the events that constitute the transitions are not direct subevents of the process denoted by *mover*. Nevertheless, the fact verbs of movement allow causative/non-causative alternations proves that they denote transition events at some level. This way, inchoative constructions can be seen as referring to the final state of the final transition of the denoted event.

The further observation of Portuguese verbs of movement shows us that verbs that lexicalize MANNER directly related to an external cause or agent of the event, such as *empurrar* (≅ push), *arrastar* (≅ drag), *transportar* (≅ transport), do not allow non-causative constructions:

(46) a. A água empurrou a pedra. (The water pushed the rock)
a'. *A pedra empurrou(-se). (The rock pushed with the water)

b. A água arrastou a pedra. (The water dragged the rock)
b'. *A pedra arrastou(-se). (The rock dragged with the water)

c. A água transportou a pedra. (The water transported the rock)
c'. *A pedra transportou(-se). (The rock transported with the water)

The same occurs with verbs that incorporate INTENTION (see the case of *embalar* (≅ sway, animated entities, in order to get them to sleep), in (47)b), because INTENTION is also necessary related with an intentional agent, thus an external cause.

(47) a. A água balançou a pedra. (The water swayed the rock)
a'. A pedra balançou com a água. (The rock swayed with the water)

b. A mãe embalou a criança. (The mother rocked the child)
b'. *A criança embalou-se com a mãe. (The child rocked with the mother)

Note that the correlation between the prominence of an external cause or agent and the impossibility of occurring in non-causative constructions may also account for the cases of transition denoting verbs that do not allow causative/non-causative alternations. In the cases observed here, transition denoting verbs that entail INTENTION or a strong MANNER implying the action of an external cause or agent do not enter non-causative constructions.

According to our analysis of Portuguese verbs of movement, the incorporation of the considered semantic elements is usually reflected on, or reflects, semantic and syntactic properties of the lexical items. However, and when it comes to MANNER lexicalization, this generalization on its own does not suffice to predict which verbs occur in non-causative constructions since there are also MANNER denoting verbs that occur in this construction:

(48) a. A água balançou a pedra. (The water swayed the rock)
a'. A pedra balançou com a água. (The rock sawyed with the water)

b. O vento abanou a árvore. (The wind shook the tree)
b'. A árvore abanou com o vento. (The tree shook with the wind)

c. O vento rolou/rebolou a bola. (The wind rolled the ball)
c'. A bola rolou/rebolou com o vento. (The ball rolled with the wind)

On the other hand, non-causative constructions only allow the syntactic realization of cause, realized in Portuguese by PPs introduced by *com* (≅ with; indicator of cause), and do not allow the syntactic realization of agents (sentient, volitional participants that cause the event):

(49) a. *A pedra balançou com a criança. (The rock swayed with the child)

b. *A árvore abanou com a criança. (The tree shook with the child)

c. *A bola rolou/rebolou com a criança. (The ball rolled with the child)

d. *A pedra moveu-se/deslocou-se com a criança. (The rock moved with the child)

e. *A pedra recuou com a criança. (The rock moved back with the child)

It seems, thus, that the non-causative construction describes an event in which the object is affected by the event denoted by the verb, which may be caused by a non-volitional external cause whose mention is taken as irrelevant. Verbs that denote events necessarily externally caused (by a given agent or cause), such as *arrastar* (≅ drag) or *embalar* (≅ rock) cannot enter non-causative constructions.

Although the possibility of occurring in non-causative constructions is related to the semantic properties of the verbs, the decompositional analysis developed under the scope of this work does not straightforwardly predict which verbs enter and which verbs do not enter in this construction, since MANNER and INTENTION incorporation may be differently reflected in the information coded at the lexical entry level.

The following examples illustrate the lexical entries of the verbs *arrastar* (≅ drag) and *embalar* (≅ rock).

(50)

{rebolar, rolar}$_V$ [≅ *roll*]
Event type: process
Arguments: <1,2,3,4>
Subevents: e1(1,2,3,4)
Restrictions:
Head: e1
Gloss: move near the support surface in successive tumbles
Example: O empregado rebolou/rolou o barril. (The waiter rolled the barrel)
O barril rebolou/rolou. (The barrel rolled)

IS HYPONYM OF **{mover, deslocar}**$_V$ [≅ *move*]

(51)

{arrastar}V [≅ *drag*]
Event type: process
Arguments: <1, 2, 3, 4>
Subevents: e1(1,2,3,4)
Restrictions:
Head: e1
Gloss: move with effort, in contact with a surface that offers significant resistance to movement
Example: O homem arrastou a mesa. (the man dragged the table)

IS HYPONYM OF → **{mover, deslocar}**V [≅ *move*]
INCORPORATES → 4 **{com}**Prep [≅ *with*; indicator of manner] conj 1:1
INCORPORATES → 4 **{esforço}**N [≅ *effort*] conj 2:1

(52)

{embalar}V [≅ *rock*]
Event type: process
Arguments: <1, 2, 3>
Subevents: e1(1,2,3)
Restrictions:
Head: e1
Gloss: sway, persons, in order to get them to sleep
Example: A ama embalou a criança. (the nanny rocked the child)

IS HYPONYM OF → **{abanar}**V [≅ *sway*]
SELECTS → 1 **{pessoa}**N [≅ person]
SELECTS → 2 **{pessoa}**N [≅ *person*]
INCORPORATES → 3 **{para}**Prep [≅ to, in order to; indicator of intention] conj 1:1
INCORPORATES → 3 **{adormecer}**V [≅ *fall asleep*] conj 2:1

In the case of the verb *embalar* (≅ rock), in (52), the incorporation of a third argument introduced by the preposition *para* (≅ indicator of intention) may directly establish the lexical semantic properties that condition the occurrence of the verbs in this node in non-causative sentences, since the concept of INTENTION is directly represented. However, and as it is possible to see in (50) and (51), the interpretation of the MANNER component incorporated in these verbs meaning as conditioning the entailment of a necessary external cause of the event is determined by the interpretation of the MANNER denoting argument introduced by *com* ({esforço}$_N$ (≅ effort), that may be interpreted as denoting a concept that modifies the action of a given entity), in the case of *arrastar* (≅ drag), whereas in the case of *rebolar* (≅ roll) the MANNER semantic component does not correspond to any lexicalized concept.

The extraction of the information relevant for determining the occurrence in causative/inchoative alternation constructions, although related to the semantic properties of the denoted events, requires an inference device, which may profit from the representation strategies presented within the scope of this work.

8.4 The -*SE* distribution

As discussed in the introduction of this work, one of the phenomena non consistent within the class of Portuguese verbs of movement is the clitic *SE* distribution in middle and non-causative constructions.

In Portuguese, apart from reflexive verbs, there are three constructions in which the -*se* occurs: passives with -*se*, middle constructions and non-causative constructions. The table below summarizes the characteristic properties that are traditionally considered as differentiating these three constructions.

Middle construction	**Non-causative construction**	**Passive with SE**
ascribes a generic state or property	does not ascribe a generic state or property	does not ascribe a generic state or property
implies an evaluative interpretation (expressed by an adverbial modifier, for instance)	–	–
the syntactic subject has generic value, interpreted as "this type of...	the syntactic subject does not have generic value	–
the verb denotes present time	–	–
can occur without -*se*	can occur without -*se*	requires -*se*

Table 1: Middle, non-causative and passive with -se constructions

In the passive with -*se* constructions, the presence of the particle is obligatory. However in what concerns middle and non-causative constructions there are three possibilities:

i) verbs that require the presence of -*se*;

(53) a. Estas portas movem-se/*movem facilmente. (*middle construction*)
(These doors move-SE/move easily)

b. A porta moveu-se/*moveu (com o vento). (*non-causative construction*)
(The door moved-SE/moved (with the wind))

ii) verbs that do not accept the presence of -*se*;

(54) a. O barco *recuou-se/recuou (com o vento). (*non-causative construction*)
(The ship moved back-SE/moved back (with the wind))

iii) verbs that occur with or without -*se*.

(55) a. Estas portas abrem-se/abrem facilmente. (*middle construction*)
(These doors open-SE/open easily)

b. A porta abriu-se/abriu (com o vento). (*non-causative construction*)
(The door opened-SE/opened (with the wind))

Let us consider first the contexts in which the clitic is optional in middle constructions. The analysis of these contexts seems to indicate that the presence of the -*se* induces the interpretation of the involvement of an external actor, an agent, in the denoted event. For instance, the sentences in (56) and (57), a and b, seem to convey a slight meaning difference:

(56) a. Estas portas abrem facilmente. (These doors open easily)
(≅ these doors have as generic property the fact of opening easily)

b. Estas portas abrem-se facilmente (These doors open-SE easily)
(≅ these doors have as generic property the fact of **being easily opened**.)

(57) a. Este líquido ferve rapidamente. (This water boils quickly)
(≅ this liquid has as generic property the fact of boiling quickly.)

b. Este líquido ferve-se rapidamente (This water boils-SE quickly)
(≅ this liquid has as generic property the fact of **being quickly boiled**.)

Should this be the case, the presence of -*se* can be related to the fact that, in Portuguese, verbs that denote events with non-affected objects typically require the presence of the clitic in middle constructions, indicating that there is necessarily the involvement of an external actor, typically an agent.

(58) a. Livros policiais lêem-se/*lêem facilmente.
(crime novels read-SE/read easily)

b. Modelos bonitos vêem-se/*vêem com agrado.
(Good looking models see-SE/see with pleasure)

c. Livros policiais vendem-se/vendem bem.
(crime novels sell-SE/sell well)

However, the verb *vender* (sell), in (58), does not require the clitic. This contrast can be related to the fact that *ler* (read) and *ver* (see) impose stronger semantic constraints on the subject, namely that it refers to human or animal entities, not easily allowing the common extension of these properties to other type of entities, as opposed to what happens with the verb *vender* (sell), as the sentences below exemplify:

(59) a. O João/#a empresa leu o livro.
(John/the company read the book)

b. O João/o gato/#a empresa viu o livro.
(John/the cat/*the company)

c. O João/a empresa vendeu o livro.
(John/the company sold the book)

Also, the presence of *-se* in non-causative constructions with verbs that do not allow the clitic, such as the verb *ferver* (≅ boil), in (60), seem to result in a passive with -*se* construction.

(60) a. A água ferveu. (The water boiled)
(≅ the water boiled)

b. A água ferveu-se. (The water boiled-SE)
(≅ the water was boiled.)

As discussed in chapter 2, the presence of the -*se* may also distinguish non-causative senses from reflex senses. The verbs {rebolar, rolar}$_V$ (≅ roll) and the verbs {rebolar-se, rebolar, rolar}$_V$ (≅ roll oneself) constitute such a contrastive pair. In (61)a we have a non-causative and non-intentional denoting sentence in which the two-place verbs {rebolar, rolar}$_V$ occur, without -*se*. In (61)b, we have a reflex event denoting verb (i.e. an event in which the agent and the patient/theme participants are co-referent) that cannot co-occur with cause denoting PPs, or in contexts forcing a non-causative reading. Naturally, non-causative construction cannot occur with intention denoting PPs ((62)b) while reflex constructions can ((62)a).

(61) a. O soldado caiu do tanque <u>já inconsciente</u> e rebolou/rolou pela encosta <u>por causa de/com o peso</u>. (The soldier fell from the tank <u>already unconscious</u> and rolled through the hill <u>with/because of the weight</u>)

b. *O soldado caiu do tanque <u>já inconsciente</u> e rebolou-se/rolou-se pela encosta <u>por causa de/com o peso</u>. (The soldier fell from the tank <u>already unconscious</u> and rolled-SE through the hill <u>with/because of the weight</u>).

(62) a. O soldado rebolou-se/rebolou/rolou pela estrada <u>para evitar ser visto</u>. (The soldier rolled-SE/rolled through the street <u>to avoid being seen)</u>

b. *O soldado rebolou-se/rolou-se com/por causa do peso pela estrada para evitar ser visto. (The soldier rolled-SE/rolled through the street with/because of the weight to avoid being seen.)

These data seem to motivate the hypothesis of considering that the clitic *se* induces the interpretation of the involvement of an external actor, an agent, in the denoted event, explaining the presented contrasts: passives with -*se* necessarily entail an external cause and thus require the presence of the particle; and in middle constructions, the clitic marks the case where the involvement of an external actor in the denoted event is entailed (see contrasts illustrated in (56) and (57) above).

However, the distribution of the -*se* in non-causative constructions is not straightforwardly explained. Non-causative constructions are used to describe events in which cause is taken as irrelevant. Thus, typically, the presence of the clitic is not required. However, some verbs only occur in non-causative constructions with the -*se*:

(63) a. A caixa moveu-se (com o vento). (The box moved-SE (with the wind))
a'. *A caixa moveu (com o vento). (The box moved (with the wind))

These cases seem to illustrate some form of reflexive interpretation in the sense that the clitic marks the correlation between the agent and theme/patient participants of the event. The [-animated] property denoted by the syntactic subject forces the non-causative reading. On the contrary, the presence of the clitic with verbs that typically denote internally caused events taking [+animated] denoting subjects results in intentional reflexive senses:

(64) a. A criança aqueceu-se (com o chá): *intentional*
(The child warmed herself (with the tea))

b. A criança aqueceu (com o chá): *non intentional*
(The child warmed/the child got warm (with the tea))

Given the data presented so far, it seems to us that the distinctive characteristics of these constructions are the one presented in the table below:

Middle construction	Non-causative construction	Passive with SE
ascribes a generic state or property	does not ascribe a generic state or property	does not ascribe a generic state or property
implies an evaluative interpretation (expressed by an adverbial modifier, for instance)	–	–
the syntactic subject has generic value, interpreted as "this type of...	the syntactic subject does not have generic value	–
the verb denotes present time	–	–
can occur without *-se*: the presence of *-se* entails the involvement an external actor in the event	can occur without *–se*: the presence of *–se* forces a reflexive reading that allows non-causative interpretations with [-animated] syntactic subjects	requires *–se*
both arguments are somehow realized, thus the agent or cause cannot be expressed	only the object argument is realized, thus cause can be expressed (but not agent)	both arguments are somehow realized, thus the agent or cause cannot be expressed

Table 2: Middle, non-causative and passive with *-se* constructions, revisited

Considering these properties, the distribution of the *–se* is not entirely defined by the lexical-semantic properties of the verbs with which it may occur, since some occurrence contexts are determined by the semantics of the arguments selected by the verb ([+ animated] or [-animated]). That is, the particular properties of the argument of a given verb – not necessarily reflected on the selection properties of the verb – also impose restrictions on the occurrence of the clitic. However, and as noted with regard to causative/non-causative alternations, the modeling of a co-composition device that accounts for this distribution may profit from the representation strategies presented within the scope of this work.

8.5 Conclusions

The analysis presented here shows that the selection of measure, obstacle and region arguments is directly related to the lexical semantic properties of the verbs, although not always lexical semantic properties result in regular patterns of behavior.

Also, and in what concerns the directed motion structures, the semantic properties of verbs of movement determine their possibility of occurring in these contexts: the properties stated at the lexical entries of SOURCE, GOAL, PATH and DIRECTION denoting constituents licenses their co-occurrence with verbs of movement, whereas the information stated in the lexical entries of verbs of change of location, license and/or restrict their co-occurrence with SOURCE, GOAL, PATH and DIRECTION denoting constituents, according to the semantic elements lexicalized by the verbs and to the subcategorization properties taken into account in their argument structure. On the other hand, the occurrence with *de/desde* (≅ from/since) and *até a* (≅ to/until) expressions is directly related to the Aktionsart properties of the verbs, extendable to verbs from other semantic domains. Also, Portuguese data show that the conclusions drawned based on the co-occurrence of manner of motions verbs with the preposition *a* do not account for the occurrence of these verbs in directed motion structures.

The analysis of middle and non-causative constructions shows that the lexical semantic properties of the verbs in study are directly related to the possibility of occurrence in these constructions. Typically, verbs that lexicalize strong MANNER or INTENTION do not allow causative/non-causative alternations, although MANNER lexicalization alone does not suffice to predict which verbs enter non-causative constructions, since the licensing of non-causative contexts is also conditioned by the interpretation of the MANNER component lexicalized in the verbs. However, the extraction of the information relevant for predicting this behavior may profit from the analysis proposed within the scope of this work.

Finally, the distribution of the *–se* in middle and non-causative constructions is not also entirely defined by the lexical-semantic properties of the verbs with which it may occur, since some occurrence contexts are determined by the semantic properties of the arguments od the verbs.

However, and to conclude this chapter, it is necessary to state that the lexical semantic characterization proposed in this work, although not accounting exhaustively for the different behaviors observed, since some of these behaviors are not entirely derived from the lexical semantic properties considered, constitutes a necessary step to enable the treatment of the observed phenomena.

9. Final Remarks

This dissertation presents our contribution to the computation of verbal predicates, focusing on lexical relational networks, lexical-conceptual structure and context sensibility of Portuguese verbs of movement.

Within the field of Computational Lexical Semantics, and based on the assumption that the performance of meaning determination computational processes is largely assisted by structured and extensive lexica, providing different types of information, we focused on the analysis of a specific set of verbs from the same semantic domain in order to determine the semantic and syntactic properties of these lexical items and how this information can be related to the computation and prediction of the structures in which they occur.

The choice of verbal predicates on the one hand, and of a particular set of verbs, on the other, emerged directly from the range of phenomena to which verbal predicates are related to (syntactic and semantic constructions, argument selection and co-occurrence restrictions, and the need to take into account the information stated for other POS items), which potentiates the extension of the results achieved to other POS entries and their contribution to the overall design of the lexicon, and from the pragmatic decision to delimit the study object in such a way that sustained the conclusions reached.

The restriction to a specific semantic domain – verbs of movement being a good case study given the conceptual proximity of the elements of this class and the diversity of semantic and syntactic behavior they display – allowed us to more accurately determine the meaning of each verb, by establishing lexical-conceptual relations within a relational model of the Lexicon. Our analysis of these verbs lexical semantics is based on the meaning specificities that differentiate hyponym verbs both from their hyperonyms, and their sister nodes. The identification of the meaning components shared and those not shared by verbs of the same semantic domain motivates the determination of the relevant semantic information to be stated at the lexical entry level, as well as the structure of this information.

We put forth a proposal for a Portuguese wordnet of verbs of movement, referring the different levels of analysis that are relevant for a coherent encoding of the verbs of this class: the way lexical items are grouped in concept denoting sets and the relations established between these sets contemplate the conceptual and semantic properties of the lexical items, and the resulting organization of the lexicon, through lexical-conceptual relations, allows for determining the information that is shared.

The development of a wordnet for Portuguese verbs of movement required the definition of the top nodes of the net as well as of some other coding options, allowing testing conceptual inheritance from the higher to the lower nodes in the hierarchy. The resulting network revealed the semantic and syntactic diversity of verbs directly related, namely that semantic properties such as argument structure or Aktionsart properties are directly related to the meaning specificities of the concepts denoted, but are not straightforwardly inherited or conditioned by the semantic domain to which a given verb belongs.

Using the developed wordnet as a base, we pursued a decompositional analysis of the meaning of the Portuguese verbs of movement, focusing on the meaning specificities that differentiate each hyponym concept with regard to its hyperonym. This analysis revealed semantic incorporation patterns different from those considered to work for Romance languages by Talmy (1985, 2000a), this way corroborating the observations of Gutiérrez (2001) and Ibarretxe-Antuñano (2004), and led us to propose a new set of semantic components that are lexicalized by the verbs in study, and extendable to the analysis of verbs from other semantic domains:

(1) Semantic components lexicalized in verbal synsets

a. **MANNER**: how the event develops

b. **CAUSE**: what brings about the event

c. **INTENTION**: purpose/intended goal of the event

d. **FIGURE**: object that anchors the event

e. **GROUND**: external object with respect to which the event is put in perspective

f. **SOURCE**: initial location of the FIGURE

g. **GOAL**: final locationof the FIGURE

h. **PATH**: medium locations between the SOURCE and the GOAL

i. **DIRECTION**: way in which the motion event occurs

The decompositional analysis of the meaning of verbs allowed us to identify the semantic content specific to each hyponym, differentiating co-hyponym verbs and explaining co-hyponyms incompatibilities: co-hyponym verbs that lexicalize opposite, or otherwise incompatible, values for the same semantic element are incompatible.

The lexicalization of the semantic components considered affects the inheritance of the hyperonym properties to different degrees, namely in what concerns argument structure (number of and type of arguments, subcategorization properties and semantic restrictions on the arguments selected) and Aktionsart properties.

Among Portuguese verbs of movement, we observed the following salient patterns: the incorporation of restrictions on the semantic components SOURCE and GOAL results in an increase of the number of overt arguments, whereas the lexicalization of these components results in a decrease of the number of overt arguments, with respect to the hyperonym argument structure. The lexicalization of PATH results in adding an overt argument to the argument structure of the hyperonym verb, usually corresponding to a GROUND argument realized in object position; the incorporation of restrictions on this semantic component results in the increase of the number of overt arguments, with respect to the hyperonym argument structure, reflected in the selection of an overt argument referring the PATH of the movement and introduced by the preposition *por* (through).

Aktionsart shifts within the wordnet of Portuguese verbs of movement, i.e., hyponyms that display different Aktionsart values from those of their hyperonyms, occur with the lexicalization of GOAL and SOURCE. The lexicalization of the elements SOURCE and GOAL result in accomplishment or achievement type events, since the determination of a specific final location or position (GOAL) or initial location or position (SOURCE) establishes a limit to the event, shifting an activity type event to an accomplishment or achievement type event.

The decompositional analysis of the semantic content of verbs of movement also motivates the presence of other representation levels in which argument and event structures can be characterized. Moreover, the information within the lexicon should account for the polymorphic properties of natural language and the creative use of words, specifically context sensibility (Pustejovsky 1993, 1995, Fellbaum 1998a, 1998b, 1999, Buitelaar 1998, Fong *et al*. 2000). This way, the need of considering lexical units as informational structures becomes apparent, which, along with computational motivations is also psychologically motivated (Sag & Wasow 1999).

For these reasons, we focused on the Generative Lexicon model (GL), namely in what concerned the integration of its levels of representation on a lexical-conceptual relational lexicon – considering the information already established in this model – but also in what regards the systematicity and consistency of the informational structures used for encoding all POS. We observed the compatibility and complementary aspects of the two models, given the similarities of the information stated at the lexical level, on the one hand (part-of relations, function relations, and so on), and the fact that wordnets can function as the type lattice in which GL lexical inheritance structure is based, on the other. In the work developed here, we propose that lexical items are characterized by the information encoded at three distinct levels

– argument structure (A), event structure (E) and qualia structure (Q) – integrated in a lexical inheritance structure (I):

(2) α = <A, E, Q> $\in$ I

In order to better characterize Portuguese verbs of movement, specifically in what concerns subcategorization properties, we propose the modeling of prepositions in WordNet.PT (WN.PT) and their semantic representation at the lexical entry level within the GL model. The integration of prepositions in WN.PT follows previous research on ontological models for the representation of prepositions, namely in what concerns the concepts denoted by prepositions consensually adopted in traditional grammars and state of the art models. In a coherent and unified manner, we simultaneously account for semantically full prepositions that introduce verbal arguments but also for argument-marking prepositions.

We used these levels and elements of representation, considering also the set of semantic components proposed, in the description of the Portuguese verbs of movement, integrated in the built wordnet. This way, we achieved a complete representation of these verbs, accounting for the percolation of information within the lexicon, for the impact of semantic lexicalization in the semantic and syntactic properties of these verbs and for verbal co-hyponym compatibility.

To account for co-hyponym compatibility in nominal items, Mendes & Chaves (2001) propose an unification operation. Incompatibility is expressed considering that co-hyponym nouns are incompatible if their qualia structures do not unify: co-hyponyms are incompatible if the values for a same qualia role in their qualia structures are not equal and if one of the values is not subsumed by the other

(3) Two qualia structures do not unify if there is a role Q from two nominal qualia structures [Q=R1] and [Q=R2] where values R1 and R2 exist such that
$\neg$(R1 = R2) $\wedge$ ($\neg$subsumes(R1, R2) $\wedge$ $\neg$subsumes(R2, R1)).

Given that it is at the argument structure level that the logical arguments of a predicate are listed, and that verbal argument structure necessarily reflects the specification of the semantic elements responsible for the meaning specificities that distinguish hyperonyms from their hyponyms, and co-hyponyms from each other, we propose to account for verbal co-hyponym compatibility by indirect qualia unification, defined in the following terms:

(4) Two co-hyponym verbs are incompatible iff the arguments in their argument structures refer to incompatible co-hyponyms, i.e. if the qualia structures of these arguments do not unify.

The recursive use of available lexical structures allows the percolation of information through the hyponymy trees and enables a coherent and economic codification of the information, including significant subcategorization properties. The resulting lexical structures demonstrate that hyponymy can replace a semantic type lattice in what concerns establishing and defining semantic properties by subtyping strategies. In addition, the permeability granted by the GL model principles, in particular underspecification and co-composition, assures the necessary context flexibility to explain the diversity of syntactic behaviors (of the lexical items) directly related to lexical semantics properties.

The contribution we intended to make to the definition of a computational lexicon that models the semantic and syntactic properties of lexical items is accomplished through the integration of informational structures in a hierarchical lexicon: GL informational structures provide the structured lexical entries, and WordNet, by its nature, provides the necessary lexical hierarchy that conveys the access to other structures in the lexicon. The integration of GL representation levels, namely argument structure, qualia structure and event structure, in a wordnet, demonstrates how wordnets can support a finer-grained lexical description that provides the bases for accounting for several lexical semantic phenomena.

Without compromising the architecture of WordNet model, this enhancement strategy compensates the emphasis of this model on lexical hierarchy in detriment of the underlying semantics of lexical items and sustains the use of wordnets for building computational lexica that support generative processes to account for co-composition and the creation of meaning in context.

We propose the integration of argument structure information in WN.PT through the establishment of three new relations:

(5) SELECTS/ IS SELECTED BY relation

$\{synset\}_1$	SELECTS	$\{synset\}_2$ and
$\{synset\}_2$	IS SELECTED BY	$\{synset\}_1$

iff:

i) x $\in$ $\{synset\}_1$ and y $\in$ $\{synset\}_2$, and the syntactic realization of x requires the syntactic realization of y, or of z, z $\in$ $\{synset\}_3$ hyponym of $\{synset\}_2$.

(6) INCORPORATES/IS INCORPORATED IN relation

$\{synset\}_1$	INCORPORATES	$\{synset\}_2$ and
$\{synset\}_2$	IS INCORPORATED IN	$\{synset\}_1$

iff:

i) the concept denoted by the $\{synset\}_1$ entails the specific concept lexicalized by the $\{synset\}_2$;

ii) x $\in$ $\{synset\}_1$ and y $\in$ $\{synset\}_2$, and the co-occurrence of x and y is only licensed by subtyping or specification processes; and

iii) in case of conjoint incorporations, ii) only applies to the element with reference potential.

(7) SELECTS BY DEFAULT/IS SELECTED BY DEFAULT BY relation

$\{synset\}_1$ SELECTS BY DEFAULT $\{synset\}_2$ and

$\{synset\}_2$ IS SELECTED BY DEFAULT $\{synset\}_1$

iff:

i) the concept denoted by the $\{synset\}_1$ entails the concept denoted by the $\{synset\}_2$ but not necessarily;

ii) x $\in$ $\{synset\}_1$ and y $\in$ $\{synset\}_2$ and the co-occurrence of x and y is only licensed by subtyping or specification processes; and

iii) in case of conjoint default selections, ii) only applies to the element with reference potential.

Combined with an order feature associated to an argument list, these relations enable the expression of argument structures, subcategorization properties, as well as finer-grained descriptions of the meaning of lexical items, particularly in what concerns lexicalization phenomena.

The integration of qualia structure in wordnets is achieved by associating lexical-conceptual relations to qualia roles, without any loss of information.

The four qualia roles – Formal, Constitutive, Agentive and Telic – correspond to different aspects of the meaning of lexical items: the Constitutive role focuses on the relations between a given (semantic) object and its constituents or parts; the Formal role focuses on the stative properties that distinguish a given object within its semantic domain; the Telic role concerns the information on the function or purpose of an object; and the Agentive role focuses on the origin or causal chain involved in the bringing about of an object. In GL, the values of these roles, are filled in by semantic predicates that express the relation between the different semantic objects that define the meaning of a lexical item. In wordnets these properties can be expressed by lexical-conceptual relations established between the nodes of the net.

Keeping in mind that lexical-conceptual relations in wordnets reflect intrinsic or prototypical properties that characterize the concept lexicalized by each synset, we propose the integration of qualia information in wordnets by associating the lexical-conceptual relations available in this model to the qualia properties these relations encode, with the advantage of determining this way the semantic predicates that can be values of the qualia roles in a coherent and consistent way. Although generally this strategy only amounted to an analysis of existing lexical-conceptual relations and their association to relevant qualia roles, for the sake of homogeneity and completeness, it implied the definition of two new relations in WN.PT:

(8) RESULTS/ORIGINATES FROM relation

{synset}$_1$	RESULTS/ORIGINATES FROM	{synset}$_2$ and
{synset}$_2$	RESULTS IN/ORIGINATES	{synset}$_1$

iff:

i) {synset}$_2$ is the origin (natural or artificial) of {synset}$_1$ and {synset}$_1$ would not exist without {synset}$_2$.
ii) {synset}$_1$ results in or originates {synset}$_2$

(9) HAS AS FUNCTION/GOAL relation

{synset}$_1$	HAS AS FUNCTION/GOAL	{synset}$_2$ and
{synset}$_2$	IS FUNCTION/GOAL OF	{synset}$_1$

iff:

i) {synset}$_2$ is the function or goal that defines {synset}$_2$ and
ii) {synset}$_1$ has as defining function or goal {synset}$_2$.

Given that the IS SUBTYPE OF and HAS AS PART wordnet relations already express the generic cases of Formal and Constitutive properties, respectively, it was only necessary to define the equivalent relations to express the generic case of Agentive and Telic properties.

Since lexical-conceptual relations in wordnets already reflect intrinsic or prototypical properties that characterize the concept lexicalized by each synset, the integration of qualia role in wordnets consists of a simple and low cost process.

Concerning event structure, we propose the expression of this representation level in wordnets through a new set of features that encode the internal properties of the events: Event type, Arguments, Subevents, Restrictions and Head.

Features	**Values**
Event type	State process Transition
Arguments	<1,2,3 ..., n>
Subevents	$e_1(2,3)$, ..., $e_n(1,2)$
Restrictions	$<$, $\circ$, $<\circ$
Head	$e_{1...n}$

Table 1: Event structure encoding features

Event structure is the level of representation that concerns the internal properties of an event associated to a lexical item, and refers five internal characteristics of this event: the list of its subevents list, <E_1 =..., ..., E_n = ...>; its Aktionsart type, expressed by the values of event type; the temporal and order restrictions of its subevents; the Head subevent; and the list of arguments of the event.

Event structure comprises semantic properties that are not necessarily (or even not at all) related to external elements (being these semantic types in a type lattice or concept denoting synsets in a lexical-conceptual relational net). Contrary to argument and qualia structures, event structure cannot easily be integrated in wordnets as lexical-conceptual relations established between existent nodes, since the properties it defines are not reflected in the nodes in the lexicon.

For these reasons, we propose the integration of event structure information as additional information at the synset level, through the use of features that mirror the attributes defined. Although features convey additional information that is not visible to the system, the systematic representation of event structure information, besides providing the grounds for argument order description, enriches the descriptive power of the resource, making wordnets a richer and more structured repository of lexical semantic information that allows the extraction of argument structure and event structure of lexical items, i.e., a lexical database that can be used as a generative lexicon over which mechanisms such as co-composition, selective binding and coercion can operate.

The enhancement of wordnets proposed in this dissertation reflects the compromise between linguistic knowledge modeling, and thus concerns regarding complexity, and the representation of relevant information, useful for natural language processing tasks. This enhancement allows the thorough description of the semantic and syntactic properties of lexical items, making the relevant information available at the lexical entries' level, describing the nature of lexical meaning as well as the specific semantic contribution made by a hyponym in relation to its hyperonym, and assuring the necessary base hierarchy for a lexical default inheritance device, without compromising the general architecture of the model.

The semantic and syntactic properties considered in the lexical entries of Portuguese verbs of movement provide insights on the occurrence restrictions displayed by these these verbs in some constructions. For these reasons, we analyzed different behavior patterns of Portuguese verbs of movement in a group of contexts, and its relation to the lexical semantic properties of these verbs. We focused on the selection of arguments denoting location and GROUND occurring in object position, the expression of directed motion in Portuguese, the occurrence of verbs of movement in middle and non-causative constructions and the distribution of *–SE* in these constructions.

According to our description of Portuguese verbs of movement, there are three types of nominal expressions, not corresponding to theme objects, that can occur in object position: measure denoting expressions; obstacle/GROUND denoting expressions; and location denoting expressions. Verbs that lexicalize SOURCE & GOAL or PATH select defined GROUND objects, i.e., true arguments denoting concrete and bounded entities, syntactically expressed by NPs, and do not occur with measure denoting expressions. Also, PPs introduced by the preposition *por* (≅ through) refer to PATH, whereas PPs introduced by the preposition *em* (≅ in/at) refer to locations, and only rarely these are ambiguous, contrary to what is the case in English, where PATH and location arguments can be syntactically realized by NPs. Typically, verbs that can co-occur with PATH or location denoting expressions correspond to process denoting verbs that express undirected movement through a random and unstructured path, and are characterized by lexicalizing MANNER.

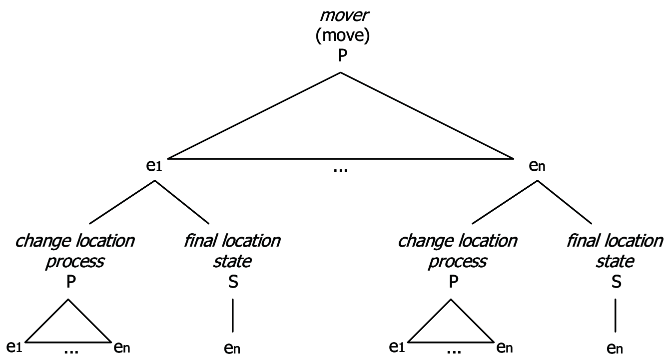
The possibility of occurring in directed motion structures, i.e. with PPs that express the SOURCE and GOAL of the movement, is frequent with verbs of movement and is directly related to the semantic and syntactic properties of the verb at stake. The information stated in the lexical entries of verbs of change of location licenses and/or restricts their co-occurrence with SOURCE, GOAL, PATH and DIRECTION denoting constituents, according to the semantic elements lexicalized by the verbs and to the subcategorization properties taken into account in their argument structure. This way, the decompositional analysis of the meaning of verbs of movement allow us to predict which verbs may occur with which constituents, without devising lexical entries for a rigid directed motion construction.

Also, Portuguese data allowed us to distinguish between expressions introduced by *de/desde* (≅ since) and *até a* (≅ until), that delimit the event denoted by the verb through temporal or spatial expressions, from expressions introduced by *de* (≅ from) and *para* (≅ to), that express SOURCE and GOAL. The occurrence of verbal predicates with the expressions *de/desde* (≅ from/since) … *até a* (≅ to/until) is licensed by the Aktionsart properties of verbs, regardless of the semantic domain they belong to. However, occurrence with PPs denoting SOURCE and GOAL is licensed by the semantic properties of the verbs, that is, by these being change of location denoting verbs.

Moreover, and regarding the expression of directed motion in Portuguese, the data show that the distribution of GOAL denoting PPs introduced by the preposition *a* (roughly corresponding to the English preposition *to* in some contexts) is conditioned by the type of movement event denoted by the verb (manner of motion vs. directed motion), but also by Aktionsart properties, since PPs introduced by *a* induce a punctual aspect interpretation of the final state of the event. This hypothesis predicts that the distribution of the preposition *a* is conditioned by the Aktionsart properties of the verbs it co-occurs with: expressions denoting goal location introduced by the preposition *a* can occur with accomplishment or achievement denoting verbs of change of location, which lexicalize GOAL and SOURCE. Also, it seems clear to us that an analysis of verbs of movement in Romance languages based solely on these verbs co-occurrence restrictions with the preposition *a* does not accurately reflects the behavior and/or properties of these verbs.

In what concerns non-causative constructions, and althought not traditionally considered in the analysis of these constructions, verbs of movement commonly occur in non-causative constructions, given the particular event structure of the events they denote.

(10) Event structure for the verb *mover* (≅ move)



The correlation between the prominence of an external cause or agent and the impossibility of occurring in non-causative constructions accounts for the possibility verbs of movement have of occurring in these constructions. Verbs that lexicalize INTENTION or a strong MANNER component implying the action of an external cause or agent do not enter non-causative constructions.

The analysis of the distribution of *–se* in middle, non-causative and passive constructions motivated our hypothesis in considering that the clitic *se* induces the interpretation of the involvement of an external actor in the denoted event, as this hypothesis explains the existing contrasts: passives with *-se* necessarily entail an external cause and thus require the presence of the clitic; in middle constructions, the clitic marks the case where the involvement of an external actor in the denoted event is entailed; and, in non-causative constructions, the clitic marks the correlation between the agent and theme/patient participants of the event, forcing the non-causative reading with [-animated] syntactic subjects.

The distribution of the *–se* in middle and non-causative constructions is, thus, not also entirely defined by the lexical-semantic properties of the verbs with which it may occur, since some occurrence contexts are determined by the semantic properties of the arguments ([-animated]/[+animated]) of the verb. However, the lexical semantic characterization proposed in this work, although not accounting exhaustively for the different behaviors observed, constitutes a necessary step to enable the treatment of the observed phenomena.

At the end of this work, and considering the achievements summarized in this chapter, we can state that the lexical semantics analysis presented shows that richer informational structures at lexical entry level are able to mirror both semantic and syntactic properties of the lexical items represented, accounting for divergent linguistic behaviors of verbs of the same class, and contributing to the computation of verbal predicates. Also, we make apparent that the modeling of lexical items of a given POS is not independent from that of others of different POS with which they may occur, which necessarily extended the scope of our analysis. Moreover, we show that modeling lexical items in the WordNet model, establishing a motivated lexical-conceptual inheritance structure, allows for an economic and adequate description of lexical items and potentiates the construction of large-scale lexical resources suitable for computational purposes.

10. References

AGIRRE, E. & D. MARTINEZ (2002), "Integrating selectional preferences in WordNet", in *Proceedings of the First International WordNet Conference*, Mysore.

AMARO, R. (2005), "Semantic Incorporation in a Portuguese WordNet of Verbs of Movement: on Aktionsart shifting", in *Proceedings of the Third International Workshop on Generative Approaches to the Lexicon*, pp. 1-9, University of Geneva, Suíça.

AMARO, R. (2006) "WordNet as a base Lexicon Model for the Computation of Verbal Predicates", in *Proceedings of The Third Global WordNet Association Conference*, Jeju Island, Coreia.

AMARO, R., R. P. CHAVES, P. MARRAFA & S. MENDES (2006), "Enriching wordnets with new Relations and with event and argument structures", in *Proceedings of CICLing 2006 – Conferences on Computational Linguistics and Intelligent Text Processing*, Mexico City, Mexico, pp. 28-40.

ASHER, N. & P. SABLAYROLLES (1996), "A Typology and Discourse Semantics for Motion Verbs and Spatial PPs in French", in Pustejovsky, J. & B. Boguraev (eds.), *Lexical Semantics: the Problem of Polysemy*, Oxford: Clarendon Press, pp. 163-209.

ATKINS, B., J. KEGL & B. LEVIN (1986), "Explicit and Implicit Information in Dictionaries", in *CSL Report 5*, Cognitive Science Laboratory, Princeton University, Princeton.

BANNARD, C. & T. BALDWIN (2003), "Distributional Models of Preposition Semantics", in *Proceedings of the ACL-SIGSEM Workshop on the Linguistic Dimensions of Prepositions and their Use in Computational Linguistics Formalisms and Applications*, Toulouse, France, pp. 169–180.

BRESNAN, J. (1982), *The Mental Representation of Grammatical Relations*, Cambridge, Massachusetts: MIT Press.

BUITELAAR, P. (1998), *CoreLex: Systematic Polysemy and Underspecification*, PhD dissertation, Brandeis University.

BUSA, F., N. Calzolari, & A. LENCI (2001), "Generative Lexicon and the SIMPLE Mode: Developing Semantic Resources for NLP," in Busa, F. & P. Bouillon *The Language of Word Meaning*, Cambridge University Press, Cambridge.

CANNESSON, E. & P. SAINT-DIZIER (2002), "Defining and representing preposition senses: A preliminary analysis", in *Proceedings of the ACL-02 Workshop on Word Sense Disambiguation: Recent Successes and Future Directions*, Philadelphia, USA, pp. 25–31.

CHAVES, R. P. (2001), "WordNet and Automated Text Summarization", in *Proceedings of the 6th Natural Language Processing Pacific Rim Symposium* - NLPRS 2001, Tokio, Japan.

CHOMSKY, N. (1981), *Lectures on Government and Binding*, Foris Publications, Dordrecht.

CINQUE, G. (1988), "On *Si* Constructions and the Theory of *Arb*", in *Linguistic Inquiry 19*, pp. 521-581.

CLARK, P., C. FELLBAUM & J. HOBBS (2008), "Using and Extending WordNet to Support Question-Answering", in *Proc. Fourth Global WordNet Conference* (GWC´08), Hungary: University of Szeged, pp. 111-119.

CORNIPS, L. & A. HULK (1999), "Affected objects" in *Languages in Contrast* 1, pp. 191–210.

CRISTOBAL, M. (2001), "Arriving events in English and Spanish: a contrastive analysis in terms of Frame Semantics", International Computer Science Institute, http://elies.rediris.es/miscelanea/misce_5/Cristobal.pdf (last accessed in 06/03/2009).

CROFT, W. (2000), *Verbs Aspect And Argument Structure: seven chapters on causal-aspectual representation*, University of Manchester, UK and Max-Planck-Institut für evolutionäre Anthropologie, Leipzig.

CROFT, W., J. BARDDAL, W. B. HOLLMANN, V. SOTIROVA & C.TAOKA (2008), "Revising Talmy's typological classification of complex events", in *Contrastive construction grammar*, John Benjamins, Amsterdam (to appear), http://ling.uib.no/barddal/TalmyTypology-paper.pdf (last accessed in 15/04/2009).

CUNHA C. & L. CINTRA (1984), *Nova Gramática do Português Contemporâneo*, Edições João Sá da Costa, Lda, Lisboa.

DANG, H., K. KIPPER & M. PALMER (2000), "Integrating compositional semantics into a verb lexicon", in *Proceedings of the 17th Conference on Computational Linguistics*, Saarbrücken, Germany, pp. 1-9.

DAVIDSON, D. (1967), "The Logical Form of Action Sentences", in Rescher, N. (ed.), *The Logical of Decision and Action*, Pittsburgh University Press, Pittsburgh.

DE BONI, M. & S. MANANDHAR (2002), "Automated discovery of telic relations for wordnet", in *Proceedings of the First International WordNet conference*, India, http://www-users.cs.york.ac.uk/~suresh/papers/ADOTRFW.pdf (last accessed in 25/02/2006).

DORR, B. J. (1997), "Large-scale dictionary construction for foreign language tutoring and interlingual machine translation", in *Machine Translation*, 12 (4): 271–322.

DOWTY, D. (2000), "'The Garden Swarms with Bees' and the Fallacy of 'Argument Alternation'", in Ravin, Y. & C. Leacock (eds.), *Polysemy: Theoretical and Computational Approaches*, Oxford University Press, Oxford, pp.111-127.

EVANS, V. & M. GREEN (2006), *Cognitive Linguistics: An Introduction*, Edinburgh University Press, Edinburgh.

FABRE C. & P. SÉBILLOT (1999), "Semantic interpretation of binominal sequences and information retrieval", in *Proceedings of International ICSC Congress on Computational Intelligence: Methods and Applications*. CIMA'99, Symposium on Advances in Intelligent Data Analysis, AIDA'99, New York, http://www.irisa.fr/texmex/Papers/1999/sebillotFS99.pdf. (last accessed in 06/03/2006)

FAGAN, S. (1992),*The syntax and semantics of middle constructions. A study with special reference to German*, Cambridge University Press, Cambridge.

FELLBAUM, C. (1998), "A Semantic Network of English Verbs", in Fellbaum, C. (ed.), *WordNet. An Electronic Lexical Database*, MA: The MIT Press, pp. 69-104.

FELLBAUM, C. (1998), "A semantic network of English: the mother of all wordnets", in Vossen, P. (ed.) *EuroWordNet: a multilingual database with lexical semantic networks*, Dordrecht: Kluwer Academic Publishers, pp. 137-148.

FELLBAUM, C. (ed.) (1998), *WordNet. An Electronic Lexical Database*, MA: The MIT Press.

FELLBAUM, C. (1999), "The Organization of Verbs ans Verb Concepts in a Semantic Net", in Saint-Dizier, P. (ed.), *Predicative Forms in Natural Language and in Lexical Knowledge Bases*, 93-109, Kluwer Academic Publishers, Netherlands.

FELLBAUM, C. (2003), "VP idioms in a lexical ontology", in *DGFS2003 Proceedings*, Munique.

FELLBAUM, C. & A. ZRIBI-HERTZ (1989), "La construction moyenne en fran_cais et en anglais: étude de syntaxe et de sémantique comparée", in *Recherches linguistiques à Vincennes*, vol. 18, pp. 19-57.

FELLBAUM, C. & G. MILLER (1990), "Folk psychology or semantic entailment? A reply to Rips and Conrad", in *The Psychological Review*, 97, pp. 565-570.

FÉREZ, P. C. (2008), *Motion In English And Spanish: A Perspective From Cognitive Linguistics, Typology And Psycholinguistics*, PhD dissertation, Universidad de Murcia.

FILLMORE, C. J., C. WOOTERS & C. F. BAKER (2000), *Building a Large Lexical Databank Which Provides Deep Semantics*, ICSI, University of California at Berkeley.

FERNÁNDEZ, A., M. MARTÍ, G. VÁZQUEZ & I. CASTELLÓN (1999), *On the concept of diathesis alternations as semantic oppositions*, http://www.cis.upenn.edu/siglex99 (last accessed in 06/03/2009).

FONG, S., C. FELLBAUM & D. LEBEAUX (2000), "Semantic Templates and Transitivity Alternations in the Lexicon", in *Proceedings of Conférence TALN 2000*, Lausanne, http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=F0F9BED39CC1C1562888A8BC5EDBF47B?doi=10.1.1.6.2870&rep=rep1&type=pdf (last accessed in 06/03/2009).

FONG, S. & C. FELLBAUM (2003), "Obstacles and regions", in *DGFS2003 Proceedings*, Munique, http://www.cip.ifi.lmu.de/~schalley/DGfS2003/fongFellbaumPres.pdf (last accessed in 06/03/2009).

GALTON, A. (1993), "Towards an integrated logic of space, time, and motion", in *Proceedings of the 13th International Joint Conference on Artificial Intelligence (IJCAI'93)*, Chambery, France, pp. 1550-1555, http://secamlocal.ex.ac.uk/people/staff/apgalton/papers/ijcai93.pdf (last accessed in 18/07/2008).

GALTON, A. (1997), "Space, time, and movement", in Stock, O. (ed.), *Spatial and temporal reasoning*, Kluwer Academic Publishers, pp. 321-352.

GAWRON, J. M (1985), "A Parsimonious Semantics for Prepositions and Cause", in *Papers from the Parasession on Causatives and Agentivity*, *Chicago Lingusitics Society* 21, 2 pp. 32-47.

GAWRON, J. M (1986), "Situations and Prepositions", in *Linguistics and Philosophy* 9, pp. 427-476.

GENNARI, S., S. SLOMAN, B. Malt & W. TECUMSEH FITCH (2001), "Motion events in language and cognition", in *Cognition*, vol. 83, number 1, Elsevier, pp. 49-79, http://homepage.mac.com/silvia_gennari/pubs/assets/motion_events.pdf (last accessed in 06/03/2009).

GODOY, L. (2008), "Preposições e os verbos transitivos indiretos: interface sintaxe-semântica lexical", in *Revista da ABRALIN*, vol. VII, pp. 49-68, http://www.abralin.org/revista/rv7n1/03-Luisa-Godoy.pdf (last accessed in 23/02/2009).

GOLDBERG, A. (1995), *Constructions: A Construction Grammar Approach to Argument Structure.* The University Chicago Press, Chicago.

GOLDBERG A. & R. JACKENDOFF (2004), "The English Resultative as a Family of Constructions", in *Language,* 80, pp. 532-568.

GRIMSHAW, J. (1990), *Argument Structure*, MA: The MIT Press, Cambridge.

GUTIÉRREZ, J. (2001), *Directed Motion in English and Spanish*, Estudios de Lingüística Española, vol. 11, http://elies.rediris.es/elies11 (last accessed in 06/03/2009).

HANKS, P. (2003), "Lexicography", in R. Mitkov (ed.) *The Oxford Handbook of Computational Linguistics*, Oxford: Oxford University Press, pp. 48-69.

HARABAGIU, S. & D. MOLDOVAN (1998), "Knowledge Processing on an Extended WordNet", in Fellbaum, C. (ed.), *WordNet. An Electronic Lexical Database*, MA: The MIT Press, pp. 353-378.

HASPELMATH, M. (1993), "More on the typology of inchoative/causative verb alternations", in B. Comrie & M. Polinsky (eds.) *Causatives & Transitivity*, John Benjamins, Amsterdam, pp. 88-120.

HERZOG, G. (1995), "Coping with Static and Dynamic Spatial Relations", in Amsili, P., M. Borillo & L. Vieu (eds.), *Proceedings of TSM'95, Time, Space, and Movement: Meaning and Knowledge in the Sensible World*, Château de Bonas, France, pp. 47–59, http://www.dfki.de/~flint/papers/b120.pdf (last accessed in 21/03/2006).

HOEKSTRA, T (1992), "Aspect and Theta Theory", in Roca, I. M. (ed.) *Thematic Structure: Its Role in Grammar*, Mouton de Gruyter, Berlin, pp. 145-174.

HOEKSTRA, T. & I. ROBERTS (1993), "Middles in Dutch and English", in E. Reuland & W. Abraham (eds.), *Knowledge and language vol. II: lexical and conceptual structure*, Kluwer Academic Publishers, Dordrecht, pp. 185-222.

HULK, A & L. CORNIPS (2000), "Reflexives in middles and the syntax-semantics interface", in H. Bennis & M. Everaert (eds.),*Interface Strategies*, KNAW-series, Elsevier, Amsterdam, pp. 207-222, http://www.meertens.knaw.nl/medewerkers/leonie.cornips/reflexives.pdf, (last accessed in 10/01/2007).

IBARRETXE-ANTUÑANO, I. (2004), "Motion events in Basque Narratives", in Strömqvist, S. & L. Verhoeven (eds.), *Relating Events in Narrative. Typological and Contextual Perspectives*. Hillsdale, Lawrence Erlbaum, pp. 89-112.

JACKENDOFF, R. (1983), *Semantics and Cognition*, The MIT Press, Cambridge.

JACKENDOFF, R. (1990), *Semantic Structures*, The MIT Press, Cambridge.

JAEGGLI, O. (1986), "Passive", in *Linguistic Inquiry* 17, pp. 587-622.

JENSEN, P. & J. F. NILSSON (2003), "Ontology-Based Semantics for Prepositions", in Proceedings of ACL-SIGSEM workshop: *The Linguistic Dimensions of Prepositions and their Use in Computational Linguistics Formalisms and Applications*, Toulouse, http://isvcbs.dk/~paj/toulouse2.pdf (last accessed in 06/06/2009).

JOHNSON, M. (1987) *The Body in the Mind. The Bodily Basis of Meaning, Reason and Imagination.* Chicago: Chicago University Press.

KAGEYAMA, T. (2002), *On the role of the event argument in passive, middle, and reflexive constructions*, Ms. Kwansei Gakuin University.

KILGARRIFF, A. (1993), "Inheriting Verb Alternations", in *Proceedings from the 6th European ACL*, Utrech, pp.213-221.

KIPPLE, E. & J. GURNEY (1998), "Verb Modification and the Lexicon in the Natural Language and Virtual Reality System", in *WLSS98 Second Workshop on Lexical Semantics Systems Proceedings*, Scuola Superiore, Pisa.

KOHL, K. , D. JONES, R. BERWICK & N. NOMURA (1998), "Representing Verb Alternations in WordNet", in Fellbaum, C. (ed.), *WordNet. An Electronic Lexical Database*, MA: The MIT Press, pp. 153- 178.

KRIFKA, M. (2001), "Specific Verb Classes and Alternations", http://amor.rz.hu-berlin.de/~h2816i3x/LexSemantik7.pdf (last accessed in 06/03/2009).

KUNZE, C. (1999), "Semantics of Verbs within GermaNet and EuroWordNet." in Kordoni, V. (ed.) *Proceedings on Lexical Semantics and Linking in Constraint-Based Theories. Workshop held at the 11th European Summer School in Logic, Language and Information (ESSLLI),* Utrecht University, http://folli.loria.fr/cds/1999/library/pdf/kunze.pdf (last accessed in 15/04/2009).

LAKOFF, G., (1987) *Women, Fire and Dangerous Things: What Categories Reveal about the Mind*, Chicago University Press, Chicago/London.

LANGACKER, R. (1987), *Foundations of Cognitive Grammar, Vol. I, Theoretical Prerequisites.* Stanford, Stanford University Press, CA.

LANGACKER, R. (1991), "Concept, Image and Symbol: The Cognitive Basis of Grammar", in *Cognitive Linguistic Research* 1, Mouton De Gruyter, Berlin/New York.

LAPATA, M. & C. BREW (1999), "Using subcategorization to resolve verb class ambiguity", in P. Fung & J. Zhou (eds.), *Joint SIGDAT Conference on Empirical Methods in NLP and Very Large Corpora Proceedings*, Association for Computational Linguistics, Maryland, pp. 266-274.

LEACOCK, C. & M. CHODOROW (1998), "Combining Local Context and WordNet Similarity for Word Sense Identification", in Fellbaum, C. (ed.), *WordNet. An Electronic Lexical Database*, MA: The MIT Press, pp. 265-284.

LEVIN, B. (1993), *English Verb Classes and Alternations: A Preliminary Investigation*, The University of Chicago Press.

LEVIN B. & M. RAPPAPORT HOVAC (1995), *Unaccusativity At the Syntax-Lexical Semantics Interface*, The MIT Press.

LEVINSON, S. (2003), *Space in language and cognition: explorations in Cognitive Diversity*, Cambridge University Press, Cambridge.

LITKOWSKI, K. (1978), "Models of the semantic structure of dictionaries", *American Journal of Computational Linguistics,* Mf.81*,* pp. 25-74.

LITKOWSKI, K. (2005), "Computational Lexicons and Dictionaries", *Encyclopedia of Language and Linguistics* (2nd ed.). Elsevier Publishers, Oxford, pp. 753-761, http://www.clres.com/online-papers/ell.doc (last accessed in 15/04/2009).

LOCKWOOD, K., K. FORBUS & J. USHER, J. (2005), "SpaceCase: A model of spatial preposition use", in *Proceedings of the 27th Annual Conference of the Cognitive Science Society*, Stressa, Italy, http://www.qrg.northwestern.edu/papers/files/f775-lockwood.pdf (last accessed in 23/04/2005).

LYONS, J. (1977), *Semantics,* Cambridge University Press, London.

MARRAFA, P. (1993), *Predicação Secundária e Predicados Complexos em Português. Análise e Modelização*, PhD dissertation, University of Lisbon.

MARRAFA, P. (2001), *WordNet do Português: uma base de dados de conhecimento linguístico*, Instituto Camões.

MARRAFA, P. (2002), "The Portuguese WordNet: General Architecture and Semantic Internal Relations", in *DELTA*, Brasil.

MARRAFA, P. (2003), "Rede Semântica dos Nomes em Português", in *Actas do III Coloquio Internacional de Linguística Aplicada y Sociedad*, Havana, Cuba.

MARRAFA, P. (2003), "The representation of Telic Complex Predicates in Wordnets: the case of lexical-conceptual structure deficitary verbs", in *Proceedings of Convergences03, International Conference on the Convergence of Knowledge, Culture, Language and Information Technologies*, Library of Alexandria, Alexandria, Egypt.

MARRAFA, P. (2004), "Extending WordNets to Implicit Information", in *Proceedings of LREC 2004, International Conference on Language Resources and Evaluation*, Lisbon, Portugal.

MARRAFA, P. (2005), "The representation of Telic Complex Predicates in Wordnets: the case of lexical-conceptual structure deficitary verbs", in *Research on Computing Science*, vol. 12.

MARRAFA, P., R. AMARO, R. P. CHAVES, S. LOUROSA, C. MARTINS & S. MENDES (2006), "WordNet.PT new directions", in *Proceedings of GWC'06: 3rd Global Wordnet Conference*, Jeju Island, Korea, pp. 319-321.

MARRAFA, P. & S. MENDES (2006), "Modeling Adjectives in Computational Relational Lexica", in *Proceedings of COLING/ACL 2006*, Sydney, Australia, pp. 555-562.

MARRAFA P. & S. MENDES (2007), "Using WordNet.PT for translation: disambiguation and lexical selection decisions", in *International Journal of Translation*, Vol. 19, ISSN 0940-9819, Bahri Publications.

MARRAFA, P. & H. MOURA (2005), "Portuguese Telic Causative Verbs", in *Proceedings of GL'2005 – 3rd International Workshop on Generative Approaches to the Lexicon*, University of Geneva, Switzerland, pp. 144-149.

MARRAFA, P., C. RIBEIRO & R. SANTOS (2004), "Para o Processamento da Linguagem Natural: Reutilização de Recursos Lexicais", in A*ctas da 3ª Conferência Iberoamericana en Sistemas, Cibernética e Informática*, Orlando, Florida, USA.

MARRAFA, P., C. RIBEIRO & R. SANTOS (2005), "Automatização da Geração de Dicionários Tratáveis por Máquina: Reutilização de Recursos Linguísticos", in *Revista Iberoamericana de Sistemas, Cibernética e Informática*.

MARRAFA P., C. RIBEIRO, R. SANTOS & J. CORREIA (2004), "Gathering Information from a Relational Lexical-Conceptual Database: A Natural Language Question-Answering System", in *Proceedings of the 8th World Multi-Conference on Systemics, Cybernetics and Informatics*, Orlando, Florida, USA.

McSHANE, M, S. BEALE & S. NIRENBURG (2005), *Disambiguating Homographous Prepositions and Verbal Particles In An Implemented Ontological Semantic Analyzer*, Working Paper 01-05, Institute for Language and Information Technologies University of Maryland Baltimore County, http://ilit.umbc.edu/ILIT_Working_Papers/ILIT_WP_01-05_Preps.pdf (last accessed in 13/07/2009).

MENDES, M. A. (2001), *Propriedades Sintácticas e Semânticas de Predicados Verbais com Pluralidade de Sentidos: o Caso dos Verbos psicológicos*, PhD dissertation, University of Lisbon.

MENDES, S. (2006), "Adjectives in WordNet.PT", in Soika, P., K-S. Choi, C. Fellbaum & P. Vossen (eds.), *Proceedings of The Third Global WordNet Association Conference*, Jeju Island, Korea, pp. 225-230.

MENDES, S. (2007) "Modelização dos Adjectivos em Português", in *Actas da Sessão de Estudantes Dissertando*, Comemorações do 75º aniversário do Centro de Linguística da Universidade de Lisboa, http://www.clul.ul.pt/artigos/mendes_sara.pdf, (last accessed in 10/05/2009).

MENDES, S. & R. P. CHAVES (2001), "Enriching wordnet with qualia information", in *Proceedings of the Workshop on WordNets and Other Lexical Resources*, NAACL, pp. 108-112.

MILLER, G. & P. JOHNSON-LAIRD (1976), *Language and Perception*, Harvard University Press, Cambridge, MA.

MILLER, G. (1990), "Nouns in WordNet: a lexical inheritance system", in *International Journal of Lexicography*, 3 (4), pp. 245-264.

MILLER, G. (1998), "Foreword", in C. Fellbaum (ed.), *WordNet. An Electronic Lexical Database*, The MIT Press, MA, pp. i-xxii.

MILLER G., R. BECKWITH, C. FELLBAUM, D. GROSS, K.J. MILLER (1990), "Introduction to WordNet: An On-line Lexical Database", in *International Journal of Lexicography*, vol 3, No.4, 235-244; revised version (1993), http://wordnet.princeton.edu/5papers.pdf (last accessed in 06/03/2009).

MOENS, M. (1987), *Tense, Aspect and Temporal Reference*, PhD dissertation, University of Edinburgh.

MÓIA, T. (2000), *Identifying and computing temporal locating adverbials: with a particular focus on Portuguese and English*, PhD dissertation, University of Lisbon.

NEELEMAN, A. (1997), "PP-Complements", in *Natural language and linguistic theory*, n. 15, pp. 89-137.

NIRENBURG, S. & L. LEVIN. (1991), "Syntax-Driven and Ontology-Driven Lexical Semantics" in *Proceedings of the First SIGLEX Workshop on Lexical Semantics and Knowledge Representation*. Berkeley, pp. 9-19, http://www.aclweb.org/anthology/W/W91/W91-0202.pdf (last accessed in 09/09/2009).

O'HARA T. & J. WIEBE (2003), "Classifying functional relations in factotum via word-net hypernym associations", in *Proceedings of CICLing-2003*, Mexico City, http://www.cs.nmsu.edu/~tomohara/ohara-factotum-roles-cicling03.pdf (last accessed in 06/03/2009).

O'HARA, T. & J. WIEBE (2003), "Preposition semantic classification via Treebank and FrameNet", in *Proceedings of the 7th Conference on Natural Language Learning (CoNLL-2003)*, Edmonton, Canada, pp. 79–86.

PEDERSON, B. & SØRENSEN, N. (2006), "Towards sounder taxonomies in WordNets" in *Ontolex 2006*, Genova, pp. 9-16, http://wordnet.dk/dannet/ontolexfinal.doc (last accessed in 06/03/2009).

PINKER, S. (1989), *Learnability and cognition: The acquisition of argument structure*, The MIT Press, Cambridge, MA.

PIÑON, C. (2001), *A Finer Look at Causative-Inchoative Alternation*, http://www.phil-fak.uni-duesseldorf.de/~pinon/papers/flcia.html (last accessed in 06/03/2009).

POLLARD, C. & I. SAG (1994), *Head-driven Phrase Structure Grammar*, CSLI Publications, Chicago.

PUSTEJOVSKY, J. (1991), "The Syntax of Event Structure", in *Cognition* 41, pp. 47-81.

PUSTEJOVSKY, J. (1993), "Introduction", in J. Pustejovsky (ed.), *Semantics and the Lexicon*, Studies in Linguistics and Philosophy, vol. 49, Kluwer Academic Publishers, London.

PUSTEJOVSKY, J. (1995), *The Generative Lexicon*, The MIT Press, MA.

PUSTEJOVSKY, J. (2001) "Type construction and the logic of concepts", in P. Bouillon & F. Busa (eds), *The Syntax of Word Meanings*, Cambridge University Press, http://www.cs.brandeis.edu/~jamesp/articles/type-construction.pdf (last accessed in 20/04/2009).

RADDEN, G. (2004), "The metaphor TIME AS SPACE across languages", in Baumgarten, N. *et al.* (eds.), *Übersetzen, interkulturelle Kommunikation, Spracherwerb und Sprachvermittlung—Das Leben mit Mehreren Sprachen: Festschrift für Juliane House zum 60 Geburtstag*, Aks-verlag, Bochum, pp. 226-239, http://zif.spz.tu-darmstadt.de/jg-08-2-3/docs/Radden.pdf (last accessed in 07/06/2009).

RAPPAPORT HOVAC, M. & B. LEVIN (1998), "Building Verb Meanings", in Butt, M. & W. Geuder (eds.), *The Projection of Arguments: lexical and compositional factors*, CSLI Publications, pp. 97-134.

RAPPAPORT HOVAC, M. & B. LEVIN (2002), "Change of State Verbs: Implications for Theories of Argument Projection", in *BLS* 28.

RIBEIRO, C., R. SANTOS, R. P. CHAVES & P. MARRAFA (2004), "Semi-automatic UNL Dictionary Generation using WordNet.PT", in *Proceedings of LREC 2004, International Conference on Language Resources and Evaluation*, Lisbon, Portugal.

RIBEIRO, C., R. SANTOS, J. CORREIA, R. P. CHAVES & P. MARRAFA (2004), "INQUER: A WordNet-based Question-Answering application", in *Proceedings of LREC 2004, International Conference on Language Resources and Evaluation*, Lisbon, Portugal.

RUWET, N. (1972), *Théorie Syntaxique et Syntaxe du Français*. Editions du Seuil, Paris.

SAG, I. & WASOW, T. (1999), *Syntactic Theory: A Formal Introduction*, CSLI Publications, Stanford, CA.

SAINT-DIZIER, P. (1995), "Generativity, Type Coercion and Verb Semantic Classes", in *AAAI Spring Symposium Technical Report*, pp. 145-157.

SAINT-DIZIER, P. (2005). "PrepNet: a framework for describing prepositions: preliminary investigation results", in Bunt, H., J. Geerzen & E. Thijse (eds.), *Proceedings of the 6th International Workshop on Computational Semantics (IWCS'05)*. ITK, Tilburg, pp. 25-34, www.irit.fr/~Patrick.Saint-Dizier/publi_fichier/iwcs6rev.pdf (last accessed in 20/06/2009).

SAINT-DIZIER, P. (2008), "Syntactic and Semantic Frames in PrepNet", in *International Joint Conference on Natural Language Processing* (IJCNLP 2008), ACL, Hyderabad (Inde), p. 763-768, http://www.aclweb.org/anthology-new/I/I08/I08-2106.pdf (last accessed in 07/07/2009).

SLOBIN, D. I. (2004), "The many ways to search for a frog: Linguistic typology and the expression of motion events", in Strömqvist, S. & L. Verhoeven (eds.), *Relating events in narrative: Typological and contextual perspectives in Translation*. Mahwah, Lawrence Erlbaum Associates, pp. 219-257.

SMITH, K. (2004), "Lecture 4: Aspect and Aktionsart", Linguistics 2: Semantics.

SOWA, J. (1991), "Logical Structures in the Lexicon" in Pustejovsky, J. & S. Bergler (eds.), *Lexical Semantics and Knowledge Representation*, Proceedings of the First SIGLEX Workshop, Berkeley, USA, Lecture Notes in Artificial Intelligence 627, Springer-Verlag, pp. 39-60.

TALMY, L. (1985), "Lexicalization patterns: semantic structure in lexical forms", in Shopen, T. (ed.), *Language typology and syntactic description,* Cambridge University Press, vol. III, pp. 57-149.

TALMY, L. (2000), *Toward a cognitive semantics: Vol. I: Concept Structuring System*. Cambridge, The MIT Press.

TALMY, L. (2000). *Toward a cognitive semantics: Vol. II: Typology and process in concept structuring*. Cambridge, The MIT Press.

TEIXEIRA, J. (2001), *A verbalização do espaço: modelos mentais de* frente/trás. Colecção Poliedro, University of Minho.

TENGI, R. (1998), "Design and Implementation of WordNet", in C. Fellbaum (ed.), *WordNet. An Electronic Lexical Database*, MA: The MIT Press, pp. 105-127.

TENNY, C. (1987), *Grammaticalizing aspect and affectedness*, MIT, Cambridge.

VAN VALIN, R. D. & R. LAPOLLA (1997), *Syntax: Structure, Meaning and Function*, Cambridge University Press, Cambridge.

VEALE, T. (2003), "Qualia Extraction from WordNet", in *Proceedings of the 3rd International Workshop on Creative Systems,* held as part of *IJCAI 03, the 2003 International Joint*

Conference on Artificial Intelligence, Acapulco, Mexico, http://afflatus.ucd.ie/papers/ijcai2003.pdf (last accessed in 06/03/2009).

VENDLER, Z. (1967), *Linguistics in Philosophy*, Cornell University Press, Ithaca, New York.

VERSPOOR, C. M. (1997), *Contextually-Dependent Lexical Semantics*, PhD dissertation, University of Edinburgh.

VOSSEN, P. (ed.) (1999), *EuroWordNet, A Multilingual Database with Lexical Semantic Networks,* Kluwer Academic Publishers, The Netherlands.

VOSSEN, P. (2001), "Tuning Document Based Hierarchies with Generative Principles", in *First International Workshop on Generative Approaches to the Lexicon*, http://www.vossen.info/docs/2001/TuneTreesWithQualia.pdf (last accessed in 06/03/2009).

VOSSEN, P. (ed.) (2002), "EuroWordNet General Document", version 3, University of Amsterdam, http://www.vossen.info/docs/2002/EWNGeneral.pdf (last accessed in 06/03/2009).

WILLIAMS, E. (1981), "Argument Structure and Morphology", in *The Linguistic Review*, 1, pp. 81-114.

ZLATEV, J. & P. YANGKLANG (2004), "A third way to travel: The place of Thai and serial verb languages in motion event typology", in Stromqvist, S. & L. Verhoeven (eds.), *Relating events in narrative: Typological and contextual perspectives*. Mahwah, Lawrence Erlbaum Associates, pp.159-190.

ZRIBI-HERTZ, A. (1987), "La réflexivité ergative en français moderne", in *Le français moderne*, 55, pp. 23-54.

ZUBIZARRETA M. & E. Oh (2004), *The Lexicon-Syntax Interface: The case of motion verbs.* MIT Press, http://www-rcf.usc.edu/~zubizarr/SelectedPublications.htm (last accessed in 06/03/2007).

ZWART, J-W. (1998), "Nonargument Middles in Dutch", in *Groningen Arbeiten zur Germanistischen Linguistik* 42, pp. 109-128, http://gagl.eldoc.ub.rug.nl/FILES/root/1998-42/03/GAGL-42-1998-03.pdf (last accessed in 22/03/2008)