An Open-source Lexicon for Spanish

Montserrat Marimon, Natalia Seghezzi, Núria Bel
IULA – Universitat Pompeu Fabra
Pl. de la Mercè 10-12
08002-Barcelona
{montserrat.marimon,natalia.seghezzi,nuria.bel}@upf.edu

Resumen: En este artículo presentamos el componente léxico de una gramática para el español. Nuestro objetivo es describir la información lingüística que codificamos en las entradas léxicas mediante una jerarquía de tipos con herencia múltiple de la cual se pueden extraer subconjuntos de datos necesarios para aplicaciones concretas.
Palabras clave: gramática, recursos léxicos, español.

Abstract: In this paper we describe the lexical component of a grammar for Spanish. Our aim is to depict the linguistic information we encode in the lexical entries by means of a multiple inheritance hierarchy of types from which subsets of data required for concrete applications could be extracted.
Keywords: grammar, lexical resources, Spanish.

1 Introduction

The lexical component, the repository of knowledge about the words of a particular language, plays a major role in NLP systems. The level of linguistic information that the lexicon contains –morpho-syntactic, syntactic, semantic– is determined by the application where it is used. The construction of lexical resources, however, is expensive in terms of both money and time; hence, they should be reused by more than one application.

In this paper we describe the lexical component of the *Spanish Resource Grammar* (SRG), a wide-coverage open-source[1] unification-based grammar for Spanish. Ours is a large lexicon with fine-grained information encoded by means of a multiple inheritance hierarchy of types. This paper aims to depict the linguistic information we have encoded in the lexical entries from which subsets of linguistic data required for concrete applications could be extracted.[2]

2 The Spanish Resource Grammar

The SRG is grounded in the theoretical framework of HPSG (*Head-driven Phrase Structure Grammar;* Pollard and Sag, 1994) and uses *Minimal Recursion Semantics* (MRS) for the semantic representation (Copestake et al., 2006). The SRG is implemented within the *Linguistic Knowledge Building* (LKB) system (Copestake, 2002), based on the basic components of the grammar *Matrix*, an open–source starter-kit for the development of HPSG grammars developed as part of the LinGO consortium's multilingual grammar engineering (Bender et al., 2002).

The SRG has a full coverage of close word classes and it contains about 50,000 lexical entries for open classes. The grammar also has 40 lexical rules to perform valence changing operations on lexical items and 150 structure rules to combine words and phrases into larger

[1]The SRG may be downloaded from: http://www.upf.edu/pdi/iula/montserrat.marimon/.

[2]This research was supported by the Spanish *Ministerio de Educación y Ciencia Juan de la Cierva* and *Ramon y Cajal* programmes.

constituents and to compositionally build up the semantic representation.

The SRG is part of the DELPH-IN open-source repository of linguistic resources and tools for writing (the LKB system), testing (The *[incr tsbd()]*; Oepen and Carroll, 2000) and efficiently processing HPSG grammars (the PET system; Callmeier, 2000). Further linguistic resources that are available in the DELPH-IN repository include broad-coverage grammars for English, German and Japanese as well as smaller grammars for French, Korean, Modern Greek, Norwegian and Portuguese.[3]

3 The lexicon of the SRG

The basic notion of the SRG is the *sign*. Briefly, a *sign* is a complex feature structure which conveys information about the orthographical realization of the lexical sign in STEM and syntactic and semantic information in SYNSEM. SYNSEM structures information related to the treatment of long distance dependencies in NONLOCAL and LOCAL information which includes head information that percolates up the tree structure via HEAD, subcategorization information in VAL(ENCE), whose attributes are SUBJ, COMPS, SPR and SPEC, for subject, complements, specifier and specified element, and semantic information encoded in CONT.

The MRS, encoded in the feature SYNSEM.LOCAL.CONT, is a flat semantic representation which consists of: 1) RELS - a list of semantic relations each with a "handle" (used to express scope relations) and one or more roles. Relations are classified according to the number and type of arguments; lexical relations of the same type are distinguished by the feature PRED; 2) HCONS - a set of handle constraints reflecting syntactic limitations on possible scope relations among the semantic relations; and 3) HOOK - a group of distinguished semantic attributes of a sign. These attributes are: LTOP - the local top handle, INDEX - the salient nominal instance or event variable introduced by the lexical semantic head, and XARG - the semantic index of the sign's external argument.

Each entry of the lexicon consists of a unique identifier, a lexical type (one of about 400 leaf types defined by a type hierarchy of around 5,500 types), an orthography and a semantic relation. Figure 1 shows an example.[4]

```
ejemplo_n1 := n_intr_count_le &
[ STEM < "ejemplo" >,
  SYNSEM.LKEYS.KEYREL.PRED "_ejemplo_n_rel" ].
```

Figure 1: Example of lexical entry.

In the following subsections we focus on the lexical types we have defined for open classes –main verbs, common nouns, adjectives and adverbs– and we describe the linguistic information we have encoded in each type. Due to space limits, we will only present the mostly used types. Note also that even though we will only show the most relevant LOCAL information, open class types are also defined by a set of NONLOCAL amalgamation types.

Through the type *uninflected-lexeme* we show in Figure 2, types for open classes inherit information common to all of them. This type basically identifies the HOOK's features LTOP and INDEX.

```
uninflected-lexeme := lex-item &
[ SYNSEM [
        LOCAL.CONT  [ HOOK [ LTOP #handle,
                             INDEX #ind ],
                      RELS.LIST < #key & relation &
                                  [ LBL #handle,
                                    ARG0 #ind,
                                    PRED predsort ],... > ],
        LKEYS.KEYREL #key ] ].
```

Figure 2: Basic type for open classes.

3.1 Common nouns

All common nouns are specified as taking an empty list for the valence features SUBJ and SPEC, and for MOD, since only temporal nouns and nouns in apposition may function as modifiers.[5] Common nouns take a non-empty list value for SPR; here agreement between nouns and specifiers is dealt with by identifying the INDEX of the specifier and that of the noun (#ind), which is of type *ref(erential)-ind(ex)*. Finally, common nouns get the semantic relation type *basic-noun-relation*. This

[3] See http://www.delph-in.net/.

[4] The attribute SYNSEM.LKEYS.KEYREL provides a shortcut to the semantic relation in RELS with highest scope and it is only used in the lexicon (see Figure 2).

[5] Modifying nouns are dealt with by a unary structure rule that generates a modifying nominal *sign.*

information is encoded in the type *basic-common-noun-lex*, as we show in Figure 3.

```
basic-common-noun-lex := uninflected-lexeme &
[ SYNSEM.LOCAL [
   CAT [ HEAD noun & [ MOD < > ],
         VAL [ SUBJ < >,
               SPEC < >,
               SPR < [ OPT -,
                       LOCAL.CONT.HOOK.INDEX #ind] >]],
   CONT nom-obj &
         [ HOOK.INDEX #ind & ref-ind & [ PNG.PN 3per ],
           RELS.LIST <  basic-noun-relation &
                        [ PRED nom_rel ],... > ] ].
```

Figure 3: Basic type for common nouns.

Then, lexical subtypes for nouns are basically distinguished on the basis of valence information and the *mass / countable / uncountable* distinction. This semantic classification determines the syntactic behavior of nouns w.r.t. the specifiers they may co-occur. Briefly, *countable* nouns require an specifier when they are in singular (e.g. *se sentó en *(la) silla* ((s)he sat in (the) chair)), they may co-occur with cardinals (e.g. *dos/tres sillas* two/three chairs)) and they only occur in plural with quantifying pronouns such as *poco* (few) (e.g. **poca silla/pocas sillas* (few chairs)); *uncountable* nouns cannot co-occur with partitives (e.g. **un trozo de paz* (a piece of peace)), nor with distributional quantifiers such as *cada* (each) (e.g. **cada paz* (each peace)), or with cardinals (e.g. **tres paces* (three peaces)); finally, *mass* nouns cannot co-occur with cardinals (e.g. **tres aburrimientos* (three boredoms)), but they may co-occur with partitives (e.g. *un poco de aburrimiento* (a little of boredom)).

Non-argumental common nouns; i.e. nouns taking an empty list as value for COMPS, are classified as *n_intr_count_le*, *n_intr_uncount_le* or *n_intr_mass_le*. Nouns with both a count and a mass reading (e.g. *manzana* (apple); *pastel de manzana* (apple pie) vs *tres manzanas* (three apples)) are assigned the type *n_intr_mass-or-count_le*. Besides, we have two subtypes: *n_intr_coll_le* for collective nouns (e.g. *ejército* (army)) and *n_intr_plur_le* for plural nouns (e.g. *celos* (jealousy)). Lexical semantic information is given to non-argumental nouns in the lexicon itself as value of the feature SYNSEM.LKEYS. KEYREL.ARG0.SORT. We have defined a hierarchy of types for dealing with nouns with more than one reading (e.g. *cabo*, which may be both human (sergeant) and locative (cape), takes *hum_loc* as value).

Nouns taking complements are classified into three types. Then, each type is further sub-typed according to such linguistic properties as the number and category of subcategorized for elements or the semantic relation type (i.e. the semantic roles of syntactic arguments). These three super-types distinguish:

1) quantifying nouns, which cover three subtypes: *n_pseudo-part_le* for pseudo-partitive nouns (e.g. *montón* (pile)), *n_part_le* for partitive nouns (e.g. *mayoría* (majority)) and *n_group_le* for group nouns (e.g. *grupo* (group)).
2) de-verbal nouns, which cover:
   - the type *n_subj-nom_le* for subject nominalizations (e.g. agresor (attacker)). Their syntactic argument is identified with the *arg2*. Lexical semantic information is given to subject nominalizations in the lexicon itself.
   - nouns derived from unaccusative verbs, which are typed either as *n_event-result_intr_le*, if they are intransitive (e.g. *muerte* (death)), or as *n_event-result_intr_lcomp_le*, if they take a locative complement (e.g. *salto a/hacia* (jump to/towards)). These types of nouns denote both events/processes and results (and get the lexical semantic type *abs(tract)_pro(cess)*), and they identify the syntactic argument with the *arg2*.
   - nouns denoting results derived from unergative verbs (e.g. *gruñido* (roar)) and intransitive verbs taking marked NPs (e.g. *lucha contra* (fight against)). These nouns are typed as *n_result_intr_le* and *n_result_intr_ppcomp_le*, respectively. Semantically, both classes of nouns are typed as *abs(tract)*, and identify the first argument with *arg1* and the second one with *arg2*. Marking prepositions are specified in the lexical entries.
   - nouns derived from transitive (or ditransitive) verbs denoting events/processes (e.g. *construcción* (construction), *envío* (dispatch)).

These nouns are typed as *n_trans_le*. Semantically, they are typed as *pro(cess)*, and identify the first argument with *arg1* and the second one with *arg2*.

3) Non-derived argumental nouns, such as relational nouns (e.g. *amigo* (friend)), body parts (e.g. *pierna* (leg))), de-adjectival nouns (e.g. *belleza* (beauty), *adicción a* (addiction to)) and nouns derived from measure psychological, inchoative and perception verbs (e.g. *peso* (weight), *temor* (fear)), are grouped together and distinguished according to the number and the category of the complements and countability features. Table 1 shows the subtypes we have defined for this class of nouns. The columns refer to the type name, the countability features –mass (f1), count (f2), uncount (f3)–, and subcategorized for elements: *de*(of)-marked NPs (f4), NPs marked by other prepositions than *de* (f5), finite completive clause (f6), infinitive clauses (f7) and interrogative clauses (f8). Lexical entries that belong to these types specify both their lexical semantic type and marking prepositions.

type	**f1**	**f2**	**f3**	**f4**	**f5**	**f6**	**f7**	**f 8**
n_ppde_count_le	-	+	-	+	-	-	-	-
n_ppde_uncount_le	-	-	+	+	-	-	-	-
n_ppde_mass_le	+	-	-	+	-	-	-	-
n_ppde_mass-or-count_le	+	+	-	+	-	-	-	-
n_cp_prop_count_le	-	+	-	-	-	+	-	-
n_cp_ques_count_le	-	+	-	-	-	-	-	+
n_ppde_ppcomp_count_le	-	+	-	+	+	-	-	-
n_ppde_ppcomp_uncount_le	-	-	+	+	+	-	-	-
n_ppde_ppcomp_mass_le	+	-	-	+	+	-	-	-
n_ppde_prop_fin_count_le	-	+	-	+	-	+	-	-
n_ppde_prop_fin_uncount_le	-	-	+	+	-	+	-	-
n_ppde_prop_inf_count_le	-	+	-	+	-	-	+	-
n_ppde_prop_inf_uncount_le	-	-	+	+	-	-	+	-
n_ppde_ques_count_le	-	+	-	+	-	-	-	+
n_ppde_ques_uncount_le	-	-	+	+	-	-	-	+

Table 1: Types for non-derived argumental common nouns.

The SRG has 35 types for common nouns and about 28,000 nominal entries.

3.2 Adjectives

All adjectival types inherit the information encoded in the type *basic-adjective-lex*, we show in Figure 4. This type specifies that the value for HEAD is of type *adj*, the SUBJ-list is empty, and the feature MOD takes a non-empty list whose element is a nominal *sign*. The semantic index of the element in the MOD list is identified with the external argument of the adjective (#ind). Finally, the *basic-adjective-lex* type assigns the *basic-adj-relation* type to adjectives.

```
basic-adjective-lex := uninflected-lexeme &
[ SYNSEM.LOCAL [
    CAT  [ HEAD adj &
                [ MOD < [ LOCAL [
                                CAT.HEAD noun,
                                CONT.HOOK.INDEX #xarg ]] > ],
           VAL.SUBJ < > ],
    CONT [ HOOK.XARG #xarg,
           RELS.LIST < basic-adj-relation &
                            [ PRED basic_adj_rel ],... > ] ] ].
```

Figure 4: Basic type for adjectives.

Then, adjectives in the SRG are cross-classified according to:

1) their position within the NP; i.e. whether they are pre and/or post modifiers (e.g. *el mero hecho* (the simple fact) vs *un chico listo* (a clever guy));
2) whether they are predicative or non-predicative. Predicative adjectives are in turn distinguished on the basis of the copulative verb –*ser or estar*– they may co-occur (e.g. *ser listo* (to be clever) vs *estar listo para* (to be ready for));
3) whether they are gradable or not. Gradable adjectives may be modified by intensifying adverbs (e.g. *muy guapa* (very pretty)) and may occur in

comparative and measure constructions (e.g. *más alto que Juan* (taller than Juan), *dos metros de largo* (two meters long));

4) whether they are intersective (the property applies to the noun in its absolute sense (e.g. *nieve blanca* (white snow)) or scopal (the property only applies to the modified noun (e.g. *excelente músico* (excellent musician));
5) whether they are positive (e.g. *bien* (good)), comparative (e.g. *mejor* (better)) or superlative (e.g. (el) *mejor* (best));
6) subcategorization, where we distinguish intransitive adjectives (e.g. *guapa* (pretty)), transitive adjectives taking marked NPs (e.g. *harto de la situación* (fed up with the situation)), adjectives taking finite completive clauses (e.g. *contraria a que vengan* (opposed to their coming), adjectives taking interrogative clauses (e.g. *seguro de si vendrán* (sure whether they'll come)), control adjectives (e.g. *capaz de hacerlo* (capable of doing)) and raising adjectives (e.g. *difícil de tocar* (difficult to play)).

Table 2 shows the types for adjectives in the SRG. The columns show the types and the values they take for: their position in the NP (f1), the copula verb with which they may co-occur (f2), whether they are gradable or not (f3), the type of modifier they are (f4), their degree (f5) and valence (f6); here, values are: 'i' (intranstive), 't' (transitive), 'cc' (completive clause), 'ic' (interrogative clause), 'sc' (subject control), 'oc' (object control), 'sr' (subject raising) and 'or '(object raising),

type	f1	f2	f3	f4	f5	f6
a_adv_int_le	pre	none	-	s	p	i
a_adv_event_le	pre/post	none	-	s	p	i
a_rel_prd_le	post	ser	-	i	p	i
a_rel_nprd_intr_le	post	none	-	i	p	i
a_rel_nprd_trans_le	post	none	-	i	p	t
a_rel_nprd_prop_le	post	none	-	i	p	cc
a_rel_nprd_ques_le_le	post	none	-	i	p	ic
a_qual_intr_scopal_le	pre/post	ser	+	s	p	i
a_qual_intr_ser_le	pre/post	ser	+	i	p	i
a_qual_intr_ser_pstn_le	post	ser	+	i	p	i
a_qual_intr_estar_le	post	estar	+	i	p	i
a_qual_trans_ser_le	pre/post	ser	+	i	p	t
a_qual_trans_ser_pstn_le	post	ser	+	i	p	t
a_qual_trans_estar_le	post	estar	+	i	p	t
a_qual_prop_ser_le	pre/post	ser	+	i	p	cc
a_qual_prop_estar_le	post	estar	+	i	p	cc
a_qual_ques_ser_le	pre/post	ser	+	i	p	ic
a_qual_ques_estar_le	post	estar	+	i	p	ic
a_sr_le	pre/post	ser	+	i	p	sr
a_sctrl_ser_le	pre/post	ser	+	i	p	sc
a_sctrl_estar_le	post	estar	+	i	p	sc
a_or_le	pre/post	ser	+	i	p	or
a_octrl_le	post	estar	+	i	p	oc
a_compar_le	pre	ser	+	i	c	t
a_super_le	pre/post	both	+	i	s	i

Table 2: Some types of adjectives.

Optionality is encoded in the types, which means that all types for adjectives that take complements have been doubled. Marking preposition is specified in the lexical entries.

The SRG has 44 types for adjectives and about 11,200 adjectival entries.

3.3 Adverbs

Leaving apart close classes of adverbs; i.e. deictic adverbs (e.g. *aquí* (here)), relative adverbs (e.g. *donde* (where)), interrogative adverbs (e.g. *cómo* (how),) and degree adverbs (e.g. *casi* (almost), *más* (more),...), we distinguish two types of adverbs: scopal adverbs and intersective adverbs.

As we show in Figure 5, intersective adverbs identify their *arg1* and the INDEX of the modified element, whereas scopal adverbs identify their own INDEX and that of the modified element. Scopal adverbs take the handle of the modified element as their argument, so that the modifier outscopes the head.

```
basic_intersective_adverb_lex := basic-adverb-lex &
[ SYNSEM.LOCAL [
    CAT.HEAD.MOD <[LOCAL intersective-mod &
                          [CONT.HOOK.INDEX #ind]]>,
    CONT.LKEYS.KEYREL.ARG1 #ind ] ].


basic_scopal_adverb_lex := basic-adverb-lex &
[ SYNSEM.LOCAL [
    CAT.HEAD.MOD < [ LOCAL scopal-mod &
                     [ CONT.HOOK [ LTOP #larg,
                                   INDEX #index]]]>,
    CONT [ HOOK.INDEX #index,
           HCONS <! qeq & [ HARG #harg,
                            LARG #larg ] !> ],
           LKEYS.KEYREL.ARG1 #harg ] ].
```

Figure 5: Basic types for intersective and scopal adverbs.

Through their super-type *basic-adverb-synsem*, as we show in Figure 6, both subtypes inherit information common to them, including the HEAD *adv* value, the empty-list values for both SUBJ and COMPS[6] and the identification of the external argument (XARG) of the adverb and that of the element within the MOD list (#xarg). The *basic-adverb-synsem* type assigns the *basic-adv-relation* type to adverbs.

```
basic-adverb-lex := uninflected-lexeme &
[ SYNSEM.LOCAL [
    CAT [ HEAD adv &
                [ MOD < [LOCAL.CONT.HOOK.XARG #xarg] >],
          VAL [ SUBJ < >,
                COMPS < > ] ],
   CONT [ HOOK.XARG #xarg ],
          RELS.LIST < basic-adv-relation,... > ] ] ].
```

Figure 6: Basic type for adverbs.

Scopal and intersective adverbs have subtypes specifying whether they may co-occur with degree adverbs (e.g. *muy probablemente* (very probably) vs **muy diariamente* (very daily)) and the adverb placement (e.g. **no está en casa aparentemente* ((he/she) is not at home aparently) vs *sinceramente te digo/te digo sinceramente* (frankly, I tell you)), giving the four subtypes we show in Table 3.

type	ModType	G	Position
av_s_prhd_le	scopal	-	prehead
av_s_prhd_spec_le	scopal	+	prehead
av_i_psthd_le	intersect	-	posthead
av_i_psthd_spec_le	intersect	+	posthead

Table 3: Some types of adverbs.

In addition, we have: one type for scopal adverbs that only modify sentences (e.g. *quizás* (maybe)), and two types for focus intersective adverbs which distinguish adverbs that may co-occur with degree adverbs (e.g. *muy especialmente* (very specially)) from those ones which may not (e.g. **muy solamente* (very only)).

[6]Adverbs taking complements, such as *detrás de* (after) or *antes de* (before), are treated as multi-word constructions and they get the category preposition.

The SRG has 14 types for open classes of adverbs and about 4,000 entries of adverbs.

3.4 Main verbs

Figure 2 shows *basic-main-verb-lex* type, the basic type for main verbs. This type specifies that the HEAD value of main verbs is of type *verb* and takes the negative value for the boolean feature AUX(ILIARY), an empty list for MOD(IFIES) and identifies the HEAD.TAM –tense, aspect and mood– feature with the semantic INDEX.E(VENT) (#tam). Main verbs also take an empty list as value for SPR and introduce an event semantic relation in the RELS-list.

```
basic-main-verb-lex := uninflected-lexeme &
 [ SYNSEM.LOCAL [
      CAT [ HEAD verb &  [ AUX -,
                           MOD < >,
                           TAM #tam ],
            VAL.SPR < > ],
      CONT [ HOOK.INDEX event &
                           [ E #tam ] ] ,
             RELS.LIST < event-relation &
                           [ PRED v_event_rel ], ...> ] ] ].
```

Figure 7: Basic type for main verbs.

Types for main verbs are first distinguished on the value for the SUBJ-list. Thus, we have subtypes for impersonal verbs taking an empty SUBJ-list, verbs taking a verbal subject and verbs taking a nominal subject. Then, each type is sub-typed according to the value of the COMPS-list; i.e. the number and category of elements in the COMPS-list. Also, we distinguish different types of verbs according to: 1) the lexical semantic relation type in the RELS-list; thus, for instance, intransitive verbs are classified either as unaccusative verbs, whose subject is identified with the *arg2* (e.g. *morir* (to die)), or as unergative verbs, whose subject is identified with the *arg1* (e.g. *nadar* (to swim)); 2) the verb form (finite or infinitive), mood (indicative or subjunctive) and control relation of verbal complements; 3) valence changing processes they may undergo. Optionality is encoded in the types, which means that all types dealing with optional complements have been doubled. We also have types for pronominal verbs. Semantic lexical restrictions on syntactic arguments and marking prepositions are given in the lexicon itself.

The SRG has 170 types for main verbs and about 6,600 entries for verbs. Table 4 shows the

most relevant types of verbs. The columns show the types and valence information: the category of the subject (f1; n(ominal), v(erbal), -(no subject)) and the complements they take – direct object (f2), indirect object (f3), finite completive clause (f4), infinitive (f5), interrogative clause (f6), locative complement (f7), prepositional complement (f8), marked completive clause (f9), marked infinitive (f10) and marked interrogative clause (f11).

type	**f 1**	**f 2**	**f 3**	**f 4**	**f 5**	**f 6**	**f 7**	**f 8**	**f 9**	**f 10**	**f 11**
iv_strict_intr_le	-	-	-	-	-	-	-	-	-	-	-
iv_non_pass_np_le	-	+	-	-	-	-	-	-	-	-	-
iv_cp_prop_le	-	-	-	+	+	-	-	-	-	-	-
iv_subj_prop_unacc_le	v	-	-	-	-	-	-	-	-	-	-
v_subj_prop_intr_io_le	v	-	+	-	-	-	-	-	-	-	-
v_subj_prop_intr_mrkd_np	v	-	-	-	-	-	-	+	-	-	-
v_subj_prop_trans_np_le	v	+	-	-	-	-	-	-	-	-	-
v_subj_prop_trans_prop_le	v	-	-	+	-	-	-	-	-	-	-
v_unacc_le	n	-	-	-	-	-	-	-	-	-	-
v_unacc_lcomp_le	n	-	-	-	-	-	+	-	-	-	-
v_intr_le	n	-	-	-	-	-	-	-	-	-	-
v_intr_mrkd_np_le	n	-	-	-	-	-	-	+	-	-	-
v_intr_mrkd_vinf_le	n	-	-	-	-	-	-	-	-	+	-
v_intr_mrkd_prop_fin_le	n	-	-	-	-	-	-	-	+	-	-
v_intr_mrkd_ques_le	n	-	-	-	-	-	-	-	-	-	+
v_intr_io_le	n	-	+	-	-	-	-	-	-	-	-
v_trans_np_le	n	+	-	-	-	-	-	-	-	-	-
v_trans_np_mrkd_np_le	n	+	-	-	-	-	-	+	-	-	-
v_trans_np_mrkd_vinf_le	n	+	-	-	-	-	-	-	-	+	-
v_trans_np_mrkd_prop_fin	n	+	-	-	-	-	-	-	+	-	-
v_trans_np_lcomp_le	n	+	-	-	-	-	+	-	-	-	-
v_ditrans_le	n	+	+	-	-	-	-	-	-	-	-
v_trans_prop_fin_le	n	+	-	+	-	-	-	-	-	-	-
v_sctrl_le	n	+	-	-	+	-	-	-	-	-	-
v_trans_ques_le	n	+	-	-	-	+	-	-	-	-	-
v_ditrans_prop_fin_le	n	-	+	+	-	-	-	-	-	-	-
v_ditrans_vinf_le	n	-	+	-	+	-	-	-	-	-	-
v_ditrans_ques	n	-	+	-	-	+	-	-	-	-	-
v_osr_le	n	+	-	-	-	-	-	-	-	-	-

Table 4: Some types of main verbs.

References

Emily M. Bender, Dan Flickinger and S. Oepen. 2002. The grammar Matrix. An open-source starter-kit for the rapid development of cress-linguistically consistent broad-coverage precision grammar. In proceedings of the Workshop on Grammar Engineering and Evaluation at the 19th International Conference on Computational Linguistics. Taipei, Taiwan.

Ulrich Callmeier. 2000. Pet – a platform for experimentation with efficient HPSG processing. Journal of Natural Language Engineering 6(1): Special Issue on Efficient Processing with HPSG: Methods, System, Evaluation, pages 99-108.

Ann Copestake, Dan Flickinger, Carl Pollard and Ivan A. 2006. Minimal Recursion Semantics: An Introduction. Research on Language and Computation 3.4:281-332.

Ann Copestake. 2002. Implementing Typed Features Structure Grammars. CSLI Publications.

Stephan Oepen and John Carroll. 2000. Performance Profiling for Parser Engineering. Journal of Natural Language Engineering 6(1): Special Issue on Efficient Processing with HPSG: Methods, System, Evaluation, pages 81-97.

Carl J. Pollard and Ivan A. Sag. 1994. Head-driven Phrase Structure Grammar. The University of Chicago Press, Chicago.