SemCor CORPUS

1 BASIC INFORMATION

1.1 Corpus composition

SemCor En-Ro corpus (Lupu et al., 2005; Ion, 2007) is an English-Romanian parallel corpus which was developed starting from the English SemCor (Mihalcea and Pedersen, 2003), a sense-tagged corpus created at Princeton University by the WordNet Project[1] research team, which itself was originally a subpart of the Brown balanced corpus (Kučera and Francis, 1967), containing news articles, literature, scientific and religious texts. In spite of its small dimension, SemCor has been extensively used both as training and testing data in various Word-Sense Disambiguation experiments and competitions, as word-sense annotated resources are scarce (Ng, 1997; Stetina et al., 1998; de Loupy, 1998; Mihalcea & Moldovan , 1999; Mihalcea & Moldovan, 2001).
En-Ro SemCor contains a total of 178,499 words for English and 175,603 words for Romanian (Ion, 2007).

1.2 Representation of the corpora (flat files, database, markup)

The corpus is represented in XCES format.

1.3 Character encoding

The characters are UTF8 encoded.

2 ADMINISTRATIVE INFORMATION

2.1 Contact person

Name: Dan Tufiş,
Address: Calea 13 Septembrie, no. 13, 050711
Affiliation: Research Institute for Artificial Intelligence, Romanian Academy
Position: Director
Telephone: +4021 3188103
Fax: +40 21 3188142
e-mail: tufis@racai.ro

2.2 Delivery medium (if relevant; description of the content of each piece of medium)

The resource will be uploaded on the MetaShare platform as an archive.

2.3 Copyright statement and information on IPR

[1] http://wordnet.princeton.edu/

The resource is free, license-based, for research purposes and fee license-based for commercial purposes.

3 TECHNICAL INFORMATION

3.1 Directories and files

The archive that will be uploaded on the MetaShare platform will contain 2 xml files for both English and Romanian content. Every sentence (e.g. *<xces:s id="br_a01_2_2_en">*) in any of the files has a unique identifier which corresponds to the parallel sentence in the other file.

3.2 Data structure of an entry

The corpus is structured in paragraphs, divided into sentences. Each sentence is segmented into tokens, including punctuation. Each token has a descriptor attribute containing syntactic and semantic information about its grammatical *meta-category*, *lemma*, *morpho-sintactic descriptor* (*msd) – tag*[2], *syntactic constituent membership* (NP – Noun Phrase; VP – Verb Phrase; AP – Adjectival Phrase; PP – Prepositional Phrase), associated Princeton WordNet 3.0 *word-sense* and *syntactic lexical attracted word* as a 0-based position in the current sentence.

The meta-categories are hand-made clusters created taking into consideration the empirical evidence of POS translation affinities (see the annexes of this file).

3.3 Corpora size (nmb. of tokens, MB occupied on disk)

The corpus contains 354,102 tokens (including punctuation): 178,499 for English and 175,603 for Romanian.
The Space requires on Disk is about 31Mb.

4 CONTENT INFORMATION

4.1 Type of the corpus (monolingual/multilingual, parallel/comparable, raw/annotated)

This corpus is a balanced parallel, heavily annotated corpus.

4.2 The natural language(s) of the corpus

The languages of the corpus are English and standard Romanian. The diacritics and all special characters are encoded as SGML entities.

4. 3 Domain(s)/register(s) of the corpus

[2] http://nl.ijs.si/ME/V4/msd/html/index.html

The text registers represented into the corpus are: news articles, literature, scientific and religious texts.

4.4 Annotations in the corpus (if an annotated corpus)

4.4.1 Types of annotations (paragraph mark-up, sentence mark-up, lexical mark-up, syntactic mark-up, semantic mark-up, discourse mark-up)

The corpus is annotated at paragraph, sentence, constituent group and word levels, providing morpho-lexical, syntactic and semantic information. Each token has a descriptor attribute containing syntactic and semantic information about its grammatical *meta-category*, *lemma*, *morpho-sintactic descriptor* (*msd) – tag*[3], *syntactic constituent membership* (NP – Noun Phrase; VP – Verb Phrase; AP – Adjectival Phrase; PP – Prepositional Phrase), associated Princeton WordNet 3.0 *word-sense* and *syntactic lexical attracted word* as a 0-based position in the current sentence. The meta-categories are hand-made clusters created taking into consideration the empirical evidence of POS translation affinities (see the annexes of this file).

The following example shows the detailed structure with all tags and attributes used in the annotation. For more details about the XCES format, see www.xces.org.

```
<xces:cesDoc version="0.1">
    <xces:text id="SemCor3_0_en_ro" complete="y">
        <xces:body>
            <xces:p id="p1">
                <xces:s id="br_a01_1_1_en">
                    <xces:tok base="the" msd="2+,Dd;Np#1" type="word">The</xces:tok>
                    <xces:tok base="Fulton_County_Grand_Jury" msd="8+,Np;Np#1;ili:ENG30-00031264-n;0"
                            type="word">Fulton_County_Grand_Jury</xces:tok>
                    <xces:tok base="say" msd="1+,Vmis;Vp#1;ili:ENG30-01009240-v;1"
                            type="word">said</xces:tok>
                    <xces:tok base="Friday" msd="1+,Ncns;Np#2;ili:ENG30-15164463-n;2"
                            type="word">Friday</xces:tok>
                    <xces:tok base="a" msd="21+,Ti-s;Np#3;5" type="word">an</xces:tok>
                    <xces:tok base="investigation" msd="1+,Ncns;Np#3;ili:ENG30-05800611-n;3"
                            type="word">investigation</xces:tok>
                    <xces:tok base="of" msd="5+,Sp;Pp#1;7" type="word">of</xces:tok>
                    <xces:tok base="Atlanta" msd="8+,Np;Pp#1,Np#4;ili:ENG30-09076675-n;5"
                            type="word">Atlanta</xces:tok>
                    <xces:tok base="&amp;apos;s" msd="21+,St;Pp#1,Np#4;7" type="word">&amp;apos;s</xces:tok>
                    <xces:tok base="recent" msd="1+,Afp;Pp#1,Np#4,Ap#1;ili:ENG30-01730444-s;10"
                            type="word">recent</xces:tok>
                    <xces:tok base="primary_election" msd="1+,Ncns;Pp#1,Np#4;ili:ENG30-00182571-n;3"
                            type="word">primary_election</xces:tok>
                    <xces:tok base="produce" msd="1+,Vmis;Vp#2;ili:ENG30-02141146-v;10"
                            type="word">produced</xces:tok>
                    <xces:tok base="&amp;quot;" msd="DBLQ" type="punctuation">&amp;quot;</xces:tok>
```

[3] http://nl.ijs.si/ME/V4/msd/html/index.html

```
        <xces:tok base="no" msd="22+,Dz3;Np#5;14" type="word">no</xces:tok>
        <xces:tok base="evidence" msd="1+,Ncns;Np#5;ili:ENG30-05823932-n;11"
                type="word">evidence</xces:tok>
        <xces:tok base="&amp;quot;" msd="DBLQ" type="punctuation">&amp;quot;</xces:tok>
        <xces:tok base="that" msd="31+,Cs;19" type="word">that</xces:tok>
        <xces:tok base="any" msd="22+,Di3;Np#6;18" type="word">any</xces:tok>
        <xces:tok base="irregularity" msd="1+,Ncnp;Np#6;ili:ENG30-00737188-n;19"
                type="word">irregularities</xces:tok>
        <xces:tok base="take_place" msd="1+,Vmis;Vp#3;ili:ENG30-00339934-v;14"
                type="word">took_place</xces:tok>
        <xces:tok base="." msd="PERIOD" type="punctuation">.</xces:tok>
</xces:s>
```

4.4.2 Tags (if POS/WSD/TIME/discourse/etc –tagged or parsed),

The corpus contains *morpho-syntactic* information (MSD) which has been assigned automatically with RACAI's high accuracy TTL tagger (Ion, 2007; Tufiş et al., 2008). The *grammatical meta-categories* are also marked using TTL as an unsigned integer pointing to a cluster of morpho-syntactic descriptors. The clusters were manually created based on empirical evidence of POS translation affinities (see the annexes of this file).

Another annotation is the *syntactic constituent membership* (shallow parsing info) also added by the TTL.

PWN 3.0 *WSD tags* are also present. They were manually assigned to each token representing a content word.

The last tag represents the *syntactic lexical attracted word* as a 0-based position in the current sentence. They were automatically annotated using LexPar (Ion and Barbu-Mititelu, 2006), an application using Lexical Attraction Models.

4.4.3Alignment information (if the corpus contains aligned documents: level of alignment, how it was achieved)

The alignment is encoded in the sentence ids. Sentences having the same id are reciprocal translation. The translation was performed manually by the NLP group at FII-UAIC[4]. Only 81 out of 352 original Englsih SemCor files have been translated:

- br-a01, br-a02, br-a11 to br-a15 (including),
- br-b13, br-b20,
- br-c01, br-c02, br-c04,
- br-d01 to br-d04 (including),
- br-e01, br-e02, br-e04, br-e21, br-e24, br-e29,
- br-f03, br-f10, br-f19, br-f43,
- br-g01, br-g11, br-g15,
- br-h01,
- br-j01 to br-j20 (including), br-j23, br-j37, br-j52 to

[4] Faculty of Informatics – Al. I. Cuza University, Iaşi

br-j60 (including), br-j70,
• br-k01 to br-k19 (including).

4.4.4 Attributes and their values (if annotated)

The *xces:p*, *xces:s* and *xces:tok* tags identify the level of the text under the tag: paragraph, sentence, token.

- text *id* (i.e. *SemCor3_0_en_ro*): specifies the name of the corpus and the languages it contains;
- paragraph *id* (e.g. p1545): specifies the position of the paragraph unit in corpus;
- sentence id (e.g. *br_a01_49_58_en*): represents the original English SemCor file the sentence belongs to;
- In the case of the chunk structures, the *id* attribute specifies the type of chunk and its position in the sentence (e.g. *Np#1* – Noun Phrase no. 1).

Under each <*xces:tok*> tag can be found three attributes and a word form (see italics in the example under 4.4.1):

- *base*, whose value is the dictionary form of the word form;
- *msd:* which contains: the grammatical meta Category (see the anexes), the *MSD* tag associated to the wordform, the *chunk* information, the associate PWN 3.0 *word-sense* and the *syntactic lexical attracted word* as a 0-based position in the current sentence, all separated by semicolon; (e.g. *msd="1+,Vmis;Vp#1;ili:ENG30-01009240-v;1"*)
- *type:* whose value can be either "word" or "punctuation"; (e.g. type="word">*mult*</xces:tok> or type="punctuation">.</xces:tok>)

The MSDs follows the Multext-East specifications (Erjavec, 2004)[5]. For Romanian there are 614 different MSDs (Tufiş et al. 1997). They have been slightly modified (new tags for named entities have been added) are largely described in (Tufiş and Ion, 2006)

4.5 Intended application of the corpus

The primary purpose of the corpus is to be used as training or testing data for WSD tools. Because it is a small corpus, it CANNOT be used as a reference corpus for English or Romanian. However, due to the fact that it is a balanced corpus, it can be used for building small language models for word prefixes or suffixes.

4.6 Reliability of the annotations (automatically/manually assigned) – if any

The annotations are highly reliable. The MSD tagging accuracy is at least 98%. The chunking annotation has been achieved based on a regular grammar defined over the MSD tags. The word-sense labels have been manually assigned and the syntactic lexical attracted word was annotated using a state of the art theoretical model.

[5] http://nl.ijs.si/ME/V4/msd/html/index.html

5 RELEVANT REFERENCES AND OTHER INFORMATION

1. de Loupy, C. and El-Beze, M. and Marteau, P-F. (1998). *Word Sense Disambiguation using HMM Tagger*. In Proceedings of the First International Conference on Language Resources and Evaluation, pages 1255–1258, Grenade, Spain;
2. Erjavec, T. (2004). *MULTEXT-East Version 3: Multilingual Morphosyntactic Specifications, Lexicons and Corpora*. In *Proceedings of the 4th LREC Conference*, LREC'04, Lisabona, pp. 1535 – 1538;
3. Ion, R. (2007). *Word Sense Disambiguation methods applied to English and Romanian*. Ph.D. thesis, Research Institute for Artificial Intelligence (RACAI), Romanian Academy, 153 pages;
4. Ion, R. and Barbu-Mititelu, V. (2006). *Constrained Lexical Attraction Models*. In Proceedings of the Nineteenth International Florida Artificial Intelligence Research Society Conference, pages 297–302, Menlo Park, Calif., USA. AAAI Press;
5. Kučera, H. and Francis, N.W. (1967). *Computational analysis of present-day American English*. Brown University Press, Providence, Rhode Island;
6. Lupu, M., Trandabăţ, D. and Husarciuc, M. (2005). *A Romanian SemCor Aligned to the English and Italian MultiSemCor*. In Proceedings of the Romance FrameNet Workshop and Kick-off Meeting, EuroLAN 2005, pages 20–27, Babes-Bolyai University, Cluj-Napoca, Romania;
7. Mihalcea, R. and Moldovan, D. (1999). *A method for word sense disambiguation of unrestricted text*. In Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics (ACL 1999), College Park, MA;
8. Mihalcea, R. and Moldovan, D. (2001). *A highly accurate bootstrapping algorithm for word sense disambiguation*. International Journal on Artificial Intelligence Tools, 10(1–2);
9. Mihalcea, R. and Pedersen, T. (2003). *An Evaluation Exercise for Word Alignment*. In Proceedings of the HLT-NAACL 2003 Workshop: Building and Using Parallel Texts Data Driven Machine Translation and Beyond, pages 1–10, Edmonton, Canada;
10. Ng, H.T. (1997). *Getting serious about word sense disambiguation*. In Proceedings of the ACL SIGLEX Workshop on Tagging Text with Lexical Semantics: Why, What, and How?, pages 1–7, Washington, D.C., USA;
11. Stetina, J., Kurohashi, S. and Nagao, M. (1998). *General word sense disambiguation method based on a full sentential context*. In Proceedings of the Coling-ACL'98 Workshop "Usage of WordNet in Natural Language Processing Systems", pages 1–8, Montreal;
12. Tufiș, D. and Ion, R. (2007). *Specificații pentru clasa de etichete folosite în adnotarea morfo-lexicală a limbii române*. Raport de cercetare, iunie, Institutul de Cercetări pentru inteligență artificială, 24 pages;
13. Tufiş, D., Barbu, A.M., Pătraşcu, V., Rotariu, G. and Popescu, C. (1997). *Corpora and Corpus-Based Morpho-Lexical Processing*. In Dan Tufiș, P. Andersen (eds.) "Recent Advances in Romanian Language Technology", Editura Academiei, pp. 35-56;
14. Tufiş, D., Ion, R., Ceauşu, A. and Ştefănescu, D. (2008). *RACAI's Linguistic Web Services*. In Proceedings of the 6thLREC Conference – LREC'08, Marrakech.

Annexes

English meta-categories clusters:

0	Pt3	1	Voip3s	5	S
1	Af	1	Vois	5	Sp
1	Afc	1	Von	7	Qz
1	Afp	1	Vops	8	Np
1	Afs	2	Dd	8	Np-p
1	M	2	Dd--p	8	Np-s
1	Mc	2	Dd--s	8	Npfs
1	Mc-p-d	2	Dd3	8	Npms
1	Mo	2	Dd3-p	8	Npnp
1	Mo-s-r	2	Dd3-s	8	Npns
1	Nc	2	Dd3n	8	Y
1	Nc----y	2	Dd3np	8	Yn
1	Nc-p	2	Dd3ns	10	Ds
1	Nc-p--y	2	Pd--p	10	Ds----p
1	Nc-s	2	Pd--s	10	Ds----s
1	Nc-s--y	2	Pd3	10	Ds1---p
1	Ncf	2	Pd3-p	10	Ds1---s
1	Ncf---y	2	Pd3-s	10	Ds2
1	Ncfp	2	Pd3n	10	Ds3---p
1	Ncfp--y	2	Pd3np	10	Ds3---sf
1	Ncfs	2	Pd3ns	10	Ds3---sm
1	Ncfs--y	3	Va	10	Ds3---sn
1	Ncm	3	Vacs	10	Ps
1	Ncm---y	3	Vaip	10	Ps----p
1	Ncmp	3	Vaip-p	10	Ps----s
1	Ncmp--y	3	Vaip1p	10	Ps1---p
1	Ncms	3	Vaip1s	10	Ps1---s
1	Ncms--y	3	Vaip2s	10	Ps2
1	Ncn	3	Vaip3s	10	Ps3
1	Ncn---y	3	Vais	10	Ps3---p
1	Ncnp	3	Vais-p	10	Ps3---sf
1	Ncnp--y	3	Vais1s	10	Ps3---sm
1	Ncns	3	Vais2s	12	Px
1	Ncns--y	3	Vais3s	12	Px1-p
1	Vm	3	Van	12	Px1-s
1	Vmcs	3	Vapp	12	Px2-p
1	Vmip	3	Vaps	12	Px2-s
1	Vmip-p	4	Dw	12	Px3-p
1	Vmip1s	4	Dw--------q	12	Px3-s
1	Vmip2s	4	Dw--------r	12	Px3fs
1	Vmip3s	4	Dw3-p	12	Px3ms
1	Vmis	4	Dw3-s	12	Px3ns
1	Vmis-p	4	Pw	13	Pp
1	Vmis1s	4	Pw----------q	13	Pp--pn
1	Vmis2s	4	Pw----------r	13	Pp--sn
1	Vmis3s	4	Pw---a----------q	13	Pp1
1	Vmn	4	Pw---a----------r	13	Pp1-pa
1	Vmnp	4	Pw3----------q	13	Pp1-pn
1	Vmpp	4	Pw3----------r	13	Pp1-sa
1	Vmps	4	Pw3-p	13	Pp1-sn
1	Vo	4	Pw3-s	13	Pp2
1	Voip	4	Pw3n	13	Pp2-p

13	Pp3	14	Rsp	22	Pi3
13	Pp3-pa	14	Rss	22	Pi3-p
13	Pp3-pn	15	Qn	22	Pi3-s
13	Pp3fs	16	I	22	Pi3fs
13	Pp3fsa	21	St	22	Pi3ms
13	Pp3fsn	21	Ti-s	22	Pi3n
13	Pp3ms	22	Di	22	Pi3np
13	Pp3msa	22	Di--p	22	Pi3ns
13	Pp3msn	22	Di--s	22	Pz3
13	Pp3ns	22	Di3	22	Pz3-s
14	R	22	Di3-p	22	Pz3ms
14	R-p---q	22	Di3-s	22	Pz3n
14	Rm	22	Di3n	22	Pz3ns
14	Rmc	22	Di3np	31	Cc
14	Rmp	22	Dz	31	Cc-i
14	Rmp---q	22	Dz--s	31	Cc-n
14	Rmp---r	22	Dz3	31	Cs
14	Rms	22	Dz3-s	50	DATE
14	Rsc	22	Pi	100	X

Romanian meta-categories clusters:

1	Af	1	Afpfson	1	Afsmpry
1	Af----n	1	Afpfsony	1	Afsms-n
1	Af--p-n	1	Afpfsoy	1	Afsmsoy
1	Af--poy	1	Afpfsoyy	1	Afsmsry
1	Af--pry	1	Afpfsrn	1	Afsmsvy
1	Af--s-n	1	Afpfsrny	1	M
1	Af--son	1	Afpfsry	1	Mc
1	Af--soy	1	Afpfsryy	1	Mc-p-d
1	Af--sry	1	Afpfsvn	1	Mc-p-l
1	Af--svn	1	Afpfsvy	1	Mc-p-r
1	Af--svy	1	Afpm--n	1	Mc-s-d
1	Afcfp-n	1	Afpmp-n	1	Mc-s-r
1	Afcfpoy	1	Afpmp-ny	1	Mcfp-l
1	Afcfpry	1	Afpmpoy	1	Mcfp-ln
1	Afcfson	1	Afpmpoyy	1	Mcfp-rn
1	Afcfsoy	1	Afpmpry	1	Mcfpoly
1	Afcfsrn	1	Afpmpryy	1	Mcfprln
1	Afcfsry	1	Afpms-n	1	Mcfprly
1	Afcmp-n	1	Afpms-ny	1	Mcfs-l
1	Afcmpoy	1	Afpmsoy	1	Mcfsoln
1	Afcmpry	1	Afpmsoyy	1	Mcfsoly
1	Afcms-n	1	Afpmsry	1	Mcfsrln
1	Afp	1	Afpmsryy	1	Mcfsrly
1	Afp-p-n	1	Afpmsvn	1	Mcmp-l
1	Afp-p-ny	1	Afpmsvy	1	Mcms-ln
1	Afp-poy	1	Afs	1	Mcmsoly
1	Afpf--n	1	Afsfp-n	1	Mcmsrl
1	Afpf--ny	1	Afsfpoy	1	Mcmsrly
1	Afpfp-n	1	Afsfpry	1	Mffpoly
1	Afpfp-ny	1	Afsfson	1	Mffprln
1	Afpfpon	1	Afsfsoy	1	Mffprly
1	Afpfpoy	1	Afsfsrn	1	Mffsoln
1	Afpfpoyy	1	Afsfsry	1	Mffsoly
1	Afpfpry	1	Afsmp-n	1	Mffsrln
1	Afpfpryy	1	Afsmpoy	1	Mffsrly

1	Ml-po	1	Momsolyy	1	Vm--3
1	Ml-pr	1	Momsrly	1	Vmg
1	Mlfpo	1	Momsrlyy	1	Vmg-------y
1	Mlfpr	1	Nc	1	Vmii1
1	Mlmpo	1	Nc---n	1	Vmii1-----y
1	Mlmpr	1	Nc-p-n	1	Vmii1p
1	Mmfp--n	1	Nc-poy	1	Vmii1s
1	Mmfp--ny	1	Nc-pry	1	Vmii2p
1	Mmfpo-y	1	Nc-pvy	1	Vmii2p----y
1	Mmfpo-yy	1	Nc-s	1	Vmii2s
1	Mmfpr-y	1	Nc-s-ny	1	Vmii2s----y
1	Mmfpr-yy	1	Nc-son	1	Vmii3p
1	Mmfso-n	1	Nc-soy	1	Vmii3p----y
1	Mmfso-ny	1	Nc-sry	1	Vmii3s
1	Mmfso-y	1	Ncf--n	1	Vmii3s----y
1	Mmfso-yy	1	Ncf--ny	1	Vmil1
1	Mmfsr-n	1	Ncfp-n	1	Vmil1p
1	Mmfsr-ny	1	Ncfp-ny	1	Vmil1p----y
1	Mmfsr-y	1	Ncfpoy	1	Vmil1s
1	Mmfsr-yy	1	Ncfpoyy	1	Vmil1s----y
1	Mmmpo-y	1	Ncfpry	1	Vmil2p
1	Mmmpo-yy	1	Ncfpryy	1	Vmil2p----y
1	Mmmpr-n	1	Ncfpvy	1	Vmil2s
1	Mmmpr-ny	1	Ncfs-n	1	Vmil2s----y
1	Mmmpr-y	1	Ncfson	1	Vmil3p
1	Mmmpr-yy	1	Ncfsony	1	Vmil3p----y
1	Mmmso-y	1	Ncfsoy	1	Vmil3s
1	Mmmso-yy	1	Ncfsoyy	1	Vmil3s----y
1	Mmmsr-n	1	Ncfsrn	1	Vmip1p
1	Mmmsr-ny	1	Ncfsrny	1	Vmip1p----y
1	Mmmsr-y	1	Ncfsry	1	Vmip1s
1	Mmmsr-yy	1	Ncfsryy	1	Vmip1s----y
1	Mo	1	Ncfsvn	1	Vmip2p
1	Mo---l	1	Ncfsvy	1	Vmip2p----y
1	Mo---ln	1	Ncm	1	Vmip2s
1	Mo---lny	1	Ncm--n	1	Vmip2s----y
1	Mo-s-r	1	Ncmp-n	1	Vmip3
1	Mofp-ln	1	Ncmp-ny	1	Vmip3-----y
1	Mofpoly	1	Ncmpoy	1	Vmip3p
1	Mofpolyy	1	Ncmpoyy	1	Vmip3p----y
1	Mofprly	1	Ncmpry	1	Vmip3s
1	Mofprlyy	1	Ncmpryy	1	Vmip3s----y
1	Mofs-l	1	Ncmpvy	1	Vmis1p
1	Mofsoln	1	Ncms-n	1	Vmis1p----y
1	Mofsoly	1	Ncms-ny	1	Vmis1s
1	Mofsolyy	1	Ncms-y	1	Vmis1s----y
1	Mofsrln	1	Ncmsoy	1	Vmis2p
1	Mofsrly	1	Ncmsoyy	1	Vmis2p----y
1	Mofsrlyy	1	Ncmsrn	1	Vmis2s
1	Momp-ln	1	Ncmsrny	1	Vmis2s----y
1	Mompoly	1	Ncmsry	1	Vmis3p
1	Mompolyy	1	Ncmsryy	1	Vmis3p----y
1	Momprly	1	Ncmsvn	1	Vmis3s
1	Momprlyy	1	Ncmsvny	1	Vmis3s----y
1	Moms-l	1	Ncmsvy	1	Vmm-2p
1	Moms-ln	1	Vm--1	1	Vmm-2p----y
1	Momsoly	1	Vm--2	1	Vmm-2s

1	Vmm-2s----y	2	Pd3fpo	3	Vam-2p
1	Vmnp	2	Pd3fpr	3	Vam-2s
1	Vmnp------y	2	Pd3fpr--y	3	Vanp
1	Vmp	2	Pd3fso	3	Vap--sm
1	Vmp--pf	2	Pd3fsr	3	Vap--sm---y
1	Vmp--pf---y	2	Pd3fsr--y	3	Vasp1p
1	Vmp--pm	2	Pd3mpo	3	Vasp1s
1	Vmp--pm---y	2	Pd3mpr	3	Vasp2p
1	Vmp--sf	2	Pd3mpr--y	3	Vasp2s
1	Vmp--sf---y	2	Pd3mso	3	Vasp3
1	Vmp--sm	2	Pd3msr	4	Dw3--o
1	Vmp--sm---y	2	Pd3msr--y	4	Dw3--r
1	Vmsp1p	3	Qf	4	Dw3--r---e
1	Vmsp1s	3	Va	4	Dw3-po
1	Vmsp2p	3	Va--1	4	Dw3-po---e
1	Vmsp2s	3	Va--1-----y	4	Dw3fpr
1	Vmsp3	3	Va--1p	4	Dw3fso---e
1	Vmsp3-----y	3	Va--1s	4	Dw3fsr
1	Vmsp3s	3	Va--1s----y	4	Dw3mpr
1	Vmsp3s----y	3	Va--2p	4	Dw3mso---e
2	Dd3-po	3	Va--2p----y	4	Dw3msr
2	Dd3-po---e	3	Va--2s	4	Pw3--o
2	Dd3-po---o	3	Va--2s----y	4	Pw3--r
2	Dd3-pr	3	Va--3	4	Pw3-po
2	Dd3-so	3	Va--3-----y	4	Pw3-so
2	Dd3-sr	3	Va--3p	4	Pw3fpr
2	Dd3fpo	3	Va--3p----y	4	Pw3fso
2	Dd3fpr	3	Va--3s	4	Pw3fsr
2	Dd3fpr---e	3	Va--3s----y	4	Pw3mpr
2	Dd3fpr---o	3	Vag	4	Pw3mso
2	Dd3fpr--y	3	Vag-------y	4	Pw3msr
2	Dd3fso	3	Vaii1	5	Sp
2	Dd3fso---e	3	Vaii2p	5	Spca
2	Dd3fso---o	3	Vaii2s	5	Spcg
2	Dd3fsr	3	Vaii3p	5	Spsa
2	Dd3fsr---e	3	Vaii3s	5	Spsay
2	Dd3fsr---o	3	Vail1p	5	Spsd
2	Dd3fsr--ye	3	Vail1s	5	Spsg
2	Dd3fsr--yo	3	Vail2p	5	Spsgy
2	Dd3mpo	3	Vail2s	7	Qz
2	Dd3mpr	3	Vail3p	7	Qz-y
2	Dd3mpr---e	3	Vail3s	8	Np
2	Dd3mpr---o	3	Vaip1p	8	Npfp-n
2	Dd3mpr--y	3	Vaip1s	8	Npfpoy
2	Dd3mpr--yo	3	Vaip1s----y	8	Npfpry
2	Dd3mso	3	Vaip2p	8	Npfs-n
2	Dd3mso---e	3	Vaip2s	8	Npfson
2	Dd3mso---o	3	Vaip3p	8	Npfsoy
2	Dd3msr	3	Vaip3p----y	8	Npfsrn
2	Dd3msr---e	3	Vaip3s	8	Npfsry
2	Dd3msr---o	3	Vaip3s----y	8	Npfsvy
2	Dd3msr--y	3	Vais1p	8	Npmp-n
2	Dd3msr--yo	3	Vais1s	8	Npmpoy
2	Pd3-po	3	Vais2p	8	Npmpry
2	Pd3-pr	3	Vais2s	8	Npms-n
2	Pd3-so	3	Vais3p	8	Npms-y
2	Pd3-sr	3	Vais3s	8	Npmsoy

8	Npmsry	10	Ds1fsos-y	10	Ps3---p
8	Npmsvn	10	Ds1fsrp	10	Ps3---s
8	Npmsvy	10	Ds1fsrs	10	Ps3fp-s
8	Y	10	Ds1fsrs-y	10	Ps3fsrs
8	Ya	10	Ds1mp-p	10	Ps3mp-s
8	Yn	10	Ds1mp-s	10	Ps3mprs
8	Ynfpvy	10	Ds1ms-p	10	Ps3ms-s
8	Ynfsoy	10	Ds1ms-s	12	Px3--a
8	Ynfsry	10	Ds1msrs-y	12	Px3--a--------s
8	Ynmpoy	10	Ds2---s	12	Px3--a--------w
8	Ynmpry	10	Ds2fp-p	12	Px3--a--y-----w
8	Ynmpvy	10	Ds2fp-s	12	Px3--d
8	Ynmsoy	10	Ds2fsop	12	Px3--d--------s
8	Ynmsry	10	Ds2fsos	12	Px3--d--------w
8	Ynmsvy	10	Ds2fsos-y	12	Px3--d--y-----w
8	Yp	10	Ds2fsrp	13	Pp--pa
8	Yp-p	10	Ds2fsrs	13	Pp--pd
8	Yp-so	10	Ds2fsrs-y	13	Pp--po
8	Yp-sr	10	Ds2mp-p	13	Pp--pr
8	Ypfpr	10	Ds2mp-s	13	Pp--sa
8	Ypfs	10	Ds2ms-p	13	Pp--sd
8	Ypfso	10	Ds2ms-s	13	Pp--sn
8	Ypfsr	10	Ds2msrs-y	13	Pp--so
8	Ypmpr	10	Ds3---p	13	Pp--sr
8	Ypms	10	Ds3---s	13	Pp1-pa--------w
8	Ypmso	10	Ds3fp-s	13	Pp1-pa--y-----w
8	Ypmsr	10	Ds3fsos	13	Pp1-pd--------s
8	Yr	10	Ds3fsos-y	13	Pp1-pd--------w
8	Yv	10	Ds3fsrs	13	Pp1-pd--y-----w
10	Dh--p	10	Ds3fsrs-y	13	Pp1-pr--------s
10	Dh--s	10	Ds3mp-s	13	Pp1-sa--------s
10	Dh-fso	10	Ds3ms-s	13	Pp1-sa--------w
10	Dh-fsr	10	Ds3msrs-y	13	Pp1-sa--y-----w
10	Dh1fp	10	Ps--p	13	Pp1-sd--------s
10	Dh1fs	10	Ps--s	13	Pp1-sd--------w
10	Dh1fso	10	Ps1fp-p	13	Pp1-sd--y-----w
10	Dh1fsr	10	Ps1fp-s	13	Pp1-sn--------s
10	Dh1mp	10	Ps1fsrp	13	Pp1-sr--------s
10	Dh1ms	10	Ps1fsrs	13	Pp2
10	Dh2fp	10	Ps1mp-p	13	Pp2----------s
10	Dh2fs	10	Ps1mp-s	13	Pp2-pa--------w
10	Dh2fso	10	Ps1mprp	13	Pp2-pa--y-----w
10	Dh2fsr	10	Ps1mprs	13	Pp2-pd--------s
10	Dh2mp	10	Ps1ms-p	13	Pp2-pd--------w
10	Dh2ms	10	Ps1ms-s	13	Pp2-pd--y-----w
10	Dh3fp	10	Ps2---s	13	Pp2-po--------s
10	Dh3fs	10	Ps2fp-p	13	Pp2-pr--------s
10	Dh3fso	10	Ps2fp-s	13	Pp2-s---------s
10	Dh3fsr	10	Ps2fsrp	13	Pp2-sa--------s
10	Dh3mp	10	Ps2fsrs	13	Pp2-sa--------w
10	Dh3ms	10	Ps2mp-p	13	Pp2-sa--y-----w
10	Ds--p	10	Ps2mp-s	13	Pp2-sd--------s
10	Ds--s	10	Ps2mprp	13	Pp2-sd--------w
10	Ds1fp-p	10	Ps2mprs	13	Pp2-sd--y-----w
10	Ds1fp-s	10	Ps2ms-p	13	Pp2-sn--------s
10	Ds1fsop	10	Ps2ms-s	13	Pp2-so--------s
10	Ds1fsos	10	Ps2msrs-y	13	Pp2-sr--------s

13	Pp3-p	21	T--sr	22	Di3mp
13	Pp3-p---------s	21	Td-po	22	Di3mpo
13	Pp3-pd--------w	21	Tdfpr	22	Di3mpr
13	Pp3-pd--y-----w	21	Tdfso	22	Di3mpr---e
13	Pp3-po--------s	21	Tdfsr	22	Di3ms
13	Pp3-pr--------s	21	Tdmpr	22	Di3ms----e
13	Pp3-s	21	Tdmso	22	Di3mso---e
13	Pp3-sd--------w	21	Tdmsr	22	Di3msr
13	Pp3-sd--y-----w	21	Tf-s-y	22	Di3msr---e
13	Pp3-so--------s	21	Tf-so	22	Di3msr--y
13	Pp3-sr--------s	21	Tffpoy	22	Dz3
13	Pp3fpa--------w	21	Tffpry	22	Dz3-po---e
13	Pp3fpa--y-----w	21	Tffs-y	22	Dz3fso---e
13	Pp3fpo--------s	21	Tffsoy	22	Dz3fsr---e
13	Pp3fpr--------s	21	Tfmpoy	22	Dz3mpr---e
13	Pp3fpr--y-----s	21	Tfmpry	22	Dz3mso---e
13	Pp3fs---------s	21	Tfms-y	22	Dz3msr---e
13	Pp3fsa--------w	21	Tfmsoy	22	Pi3
13	Pp3fsa--y-----w	21	Tfmsry	22	Pi3--r
13	Pp3fso--------s	21	Ti-po	22	Pi3-po
13	Pp3fsr--------s	21	Tifp-y	22	Pi3-pr
13	Pp3fsr--y-----s	21	Tifso	22	Pi3-so
13	Pp3mpa--------w	21	Tifsoy	22	Pi3-sr
13	Pp3mpa--y-----w	21	Tifsr	22	Pi3fpo
13	Pp3mpo--------s	21	Tifsry	22	Pi3fpr
13	Pp3mpr--------s	21	Timp-y	22	Pi3fso
13	Pp3mpr--y-----s	21	Timso	22	Pi3fsr
13	Pp3ms---------s	21	Timsr	22	Pi3mpo
13	Pp3msa--------w	21	Timsry	22	Pi3mpr
13	Pp3msa--y-----w	21	Ts-po	22	Pi3mso
13	Pp3mso--------s	21	Tsfp	22	Pi3msr
13	Pp3msr--------s	21	Tsfs	22	Pi3msr--y
13	Pp3msr--y-----s	21	Tsmp	22	Pz3
14	R	21	Tsms	22	Pz3-po
14	Rc	22	Di3	22	Pz3-so
14	Rgc	22	Di3------e	22	Pz3-sr
14	Rgp	22	Di3-----y	22	Pz3fpr
14	Rgpy	22	Di3--r	22	Pz3fso
14	Rgs	22	Di3--r---e	22	Pz3fsr
14	Rp	22	Di3-po	22	Pz3mpr
14	Rp-y	22	Di3-po---e	22	Pz3mso
14	Rw	22	Di3-s----e	22	Pz3msr
14	Rw-y	22	Di3-sr	31	C
14	Rz	22	Di3-sr---e	31	Cccsp
15	Qn	22	Di3-sr--y	31	Ccssp
15	Qn-y	22	Di3fp	31	Ccsspy
15	Qs	22	Di3fpo	31	Crssp
16	I	22	Di3fpr	31	Cscsp
21	T--p	22	Di3fpr---e	31	Csssp
21	T--po	22	Di3fso	31	Cssspy
21	T--pr	22	Di3fso---e	50	DATE
21	T--s	22	Di3fsr	100	X
21	T--so	22	Di3fsr---e		