

Towards a Semantic Annotation of English Television News - Building and Evaluating a Constraint Grammar FrameNet

Eckhard Bick

Institute of Language and Communication, University of Southern Denmark
Campusvej 55, DK 5230 Odense M
eckhard.bick@mail.dk

Abstract

This paper presents work on the semantic annotation of a multimodal corpus of English television news. The annotation is performed on the second-by-second-aligned transcript layer, adding verb frame categories and semantic roles on top of a morphosyntactic analysis with full dependency information. We use a rule-based method, where Constraint Grammar mapping rules are automatically generated from a syntactically anchored Framenet with about 500 frame types and 50 semantic role types. We discuss design decisions concerning the Framenet, and evaluate the coverage and performance of the pilot system on authentic news data.

1 Introduction and methodological focus

Because the communicative information contained in a multi-modal corpus is distributed across different channels, it is much more difficult to process automatically than a classical text corpus. Large multi-modal corpora, in particular, constitute a challenge to quantitative-statistical exploration or even comparative qualitative studies, because they may be too big for complete inspection, let alone extensive manual mark-up. In some types of multi-modal corpora, however, such as a film-subtitle corpus, or the television news corpus that is the object of this study, aligned transcripts or captions offer at least a partial solution, because this textual layer can be used to search the corpus and extract matching sections for closer inspection, comparison or even quantitative analysis.

The UCLA Communications Studies Archive (UCLA CSA) is a so-called monitor corpus of television news, where newscasts from a large number of channels are recorded daily in high-quality video mode, amounting to ~ 150.000 hours of recorded news, and growing by 100 programs a day (DeLiema, Steen & Turner 2012). To date only English language channels have been targeted, but the author's institution has plans to join the project with matching data for first the Scandinavian languages and German, then further European languages. This paper focuses on the linguistic annotation of the time-stamp-aligned textual layer of the corpus. Optimally, such annotation should address the following issues

- robustness in the face of spoken language data
- low error rate for basic morphosyntactic annotation
- conservation/integration of non-linguistic meta-annotation (speaker, source, time ...)
- unified tag system across languages to facilitate comparative studies
- a semantic annotation layer to support higher-level communicative studies

A well-established annotation format is the assignment of feature-attribute pairs to word tokens, expressed as tag fields and convertible to xml structures. A list of tokens with tags guarantees that all information is local and easy to filter or search, with meta-information carried along on separate lines between tokens. For the tagging/parsing task as such we have chosen the Constraint Grammar (CG) formalism (Karlsson et al. 1995, Bick 2000) which has proven robust enough for a large

variety of corpus annotation task, including speech annotation (Bick 2012). An added advantage is the fact that comparable CG systems, with similar tag sets and annotation conventions, already exist not only for English, but also for many other European languages, among them almost all Germanic and Romance languages (http://visl.sdu.dk/constraint_grammar.html). CG systems are modular, hierarchical sets of rule-based grammars targeting different linguistic levels, and while higher level analysis can be performed within the same formalism, it is a challenging task. Thus, most of the existing CG systems perform only morphosyntactic and dependency annotation, with some notable exceptions in the area of NER and semantic role annotation. The system that comes closest to the task at hand, is the Danish DanGram system which implements a framenet-based verbal classification and semantic role annotation (Bick 2011), with a category inventory of ~500 verb frames and ~50 semantic roles. For our present task, we have attempted to port lexical material from this system, and adopted its verb classification scheme, which in turn was inspired by the VerbNet classes proposed by Kipper et al. (2006), ultimately with roots in (Levine 1993), and a smaller and thus more tractable granularity than PropBank (Palmer et al. 2005). Our semantic role inventory, following the one implemented for Portuguese by (Bick 2007), is also much smaller than PropBank's, the rationale being that medium-sized category sets allow for a reasonable level of abstraction compared to the underlying lexical items, and by roughly matching the granularity of other linguistic abstractions (syntactic function inventory, PoS/morphological categories) are well suited to be integrated with the latter in automatic disambiguation systems.

2 Frame role distinctors: valency, syntactic function and semantic classes

In this vein, the distinctional backbone of our frame inventory are syntactic valency frames like <vt> (monotransitive), <vdt> (ditransitive), <to^vp-forward> (prepositional transitive with the preposition “to” and a verb-incorporated 'forward'-adverb). Each of these valency frames is assigned at least one (or more¹) verb senses, each with its

¹In 717 cases, there is more than one role combination for the same sense with the same valency, and in 11.2% multiple verb senses share the same valency frame, reflecting cases where semantic prototype or other slot filler information is needed to

own semantic frame. Depending, for instance, on the number of obligatory arguments, several valency or semantic frames may share the same verb sense, but two different verb senses will almost always differ in at least one syntactic or semantic aspect of their argument frame - guaranteeing that all senses can in principle be disambiguated exploiting a parser's argument tags and dependency links.

Currently, the EngGram FrameNet (EFN) contains 7820 verb sense for 4774 verb types, with 10.800 valency frames. For each frame, we provide a list of arguments with the following information:

1. Thematic role (Table 1)
2. Syntactic function (Table 2)
3. Morphosyntactic form (Table 4)
4. for np's, a list of typical semantic prototypes to fill the slot (Table 3)
5. An English language gloss / skeleton sentence

For about 2/3 of the frames, a best-guess link to a BFN verb sense is also provided, based on semi-automatic valency matches on EngGram-parsed BFN example sentences.

Our FrameNet uses ca. 35 core thematic roles (or case/semantic roles, Fillmore 1968), with a further 10-15 adverbial roles that are added by the semantic tagger based on syntactic context without the need of a verb frame entry (e.g. subclause function based on conjunction type). These roles are far from evenly distributed in running text. Table 1 provides some live corpus data, showing that the top 5 roles account for over half of all role taggings in running text. Note that the distribution is for all roles, not just verb frame roles, since the semantic tagger also tags some semantic relations based on nominal or adjectival valency (e.g. *abolition of X, full of Y*).

Table 1: Top 25 Semantic (Thematic) Roles

	Thematic Role	in corpus
§TH	Theme	21.91%
§ATR	Attribute	13.76%
§AG	Agent	7.07%
§LOC	Location	6.78%
§LOC-TMP	Point in time	5.44%
§PAT	Patient	4.20%
§DES	Destination/Goal	3.56%
§MES	Message	3.13%

make the distinction.

§COG	Cognizer	3.00%
§SP	Speaker	2.58%
§BEN	Beneficiary	2.48%
§ID	Identity	2.16%
§TP	Topic	1.97%
§ACT	Action	1.91%
§INC	Incorporated particle	1.91%
§EXP	Experiencer	1.73%
§RES	Result	1.49%
§STI	Stimulus	1.37%
§FIN	Purpose	1.31%
§EV	Event	1.56%
§CAU	Cause	0.98%
§ORI	Origin	0.97%
§REC	Recipient	0.80%
§EXT-TMP	Duration	0.74%
§INS	Instrument/Tool	0.62%

Other roles: §COND condition, §COM co-agent, §HOL whole, §VOC vocative, §COMP comparison, §SOA state of affairs, §MNR manner, §PART part, §VAL value, §ASS asset, §EXT extension, §PATH path, §DON donor, §CONT contents, §CONC concession, §REFL reflexive, §POSS possessor, §EFF effect, §ROLE role, §MAT material, §ROLE role, §DES-TMP temp. destination, §ORI-TMP temp. origin

Even in a case-poor language like English, we found some clear likelihood relations between thematic roles and syntactic functions (table 2). Thus, agents (§AG, §COG, §SP) are typical subject roles, while patients (§PAT), messages (§MES) and results (§RES) are typical direct object roles, and recipients (§REC) and beneficiaries (§BEN) call for dative object function.

Table 2: Major syntactic Functions with most likely roles

	Function
@SUBJ	Subject
TH (44.5%) > AG (21.3%) > COG (9.6%) > SP (8.1%) > EXP (5.2%)	
@ACC	Direct object
TH (26.9%) > PAT (11.6%) > MES > RES > STI > ACT	
@DAT	Dative object
BEN (52.8%) > REC (41.9%)	
@PIV, @SA, @OA, @ADVL	Prepositional complements
LOC (30.1%), DES (11.9%) > PAT (10.0%) > BEN > TP > ORI > ATR > COM > COMP	
@SC	Subject complem.
ATR (95.7%) > RES	

@OC	Object complem.
ATR (80.7%) > RES	

The prototypical verb frame consists of a full verb and its nominal, adverbial or subclause complements. Like most other languages, however, English has also verb incorporations that are not, in the semantical sense, complements. The simplest kind are adverb incorporates, which we mark in the valency frame, but not in the argument list:

give up - <vi-up>, *turn off* - <vt-off>

More complicated are support verb constructions, where the semantic weight and - to a certain degree - valency reside in a nominal element, typically a noun that syntactically fills a (direct or prepositional) object slot, but semantically orchestrates the other complements. While adverb incorporates are marked as such by the EngGram parser already at the syntactic level (@MV<), noun or adjective incorporates receive an ordinary syntactic tag (@ACC, @SC), but are marked with an empty §INC (incorporate) role tag at the semantic level. This is why, currently, about 14.6% of EFN valency entries include incorporated material, but the percentage of non-adverbial incorporates is still small (about a 1/10 of all incorporations).

The examples below also show the corresponding valency tags, where 'vt' means *transitive* and 'vi' *intransitive*. Governed prepositions are prefixed (e.g. <of^...>) and incorporated material is postfixed (e.g. <...-stock>)

take place - <vt-place>,
take stock of - <of^vt-stock>

Some of the constructions can be rather complex and involve dependents of an incorporated noun, prepositional phrases or a combination of particles and adverbs:

take it out on - <on^vp-it-out>,
lay in waiting - <vi-in=waiting>
call in sick - <vi-in_sick>,
take care of - <of^vp-care>

One could argue that the real frame arguments (like the noun expressing what is catered for in *take care of*) should be dependency-linked to the §INC noun *care* and the frame class marked on the latter, but for consistency and processing reasons we decided to center all dependency relations on the support verb in these cases, and also mark the

frame name on the verbal element of support constructions.

3 Frame annotation

One would assume that using argument information from our verb frame lexicon on the one hand and a functional dependency parser on the other, it should in theory be possible to annotate running text with verb senses and frame elements, simply by checking verb-argument dependencies for function and semantic class. To prove this assumption, we implemented our annotation module in the Constraint Grammar formalism, choosing this particular approach in part because that made it easier to exploit the DanGram-parser's existing CG annotation tags, but also to allow for later manual fine-tuning of rules and contextual exceptions – something that would be impossible in a probabilistic system based on machine learning. In our view, this is a clear methodological advantage, and also saved us the cost of hand-annotating a training corpus. And though the creation of EFN itself does involve manual work in its own right, we prefer this method not only because, for a linguist, it is more satisfying to express lexical knowledge directly in a lexicon format, rather than indirectly through manual corpus annotation, but also because the latter is, as a method, less effective, since it will mean repetitive work for some verbs and coverage problems for others, due to the sparse data problem inherently linked to the limited size of hand-annotated corpora.

As a first step, we adapted a converter program (framenet2cgrules.pl, Bick 2011) that turned each frame into a verb sense mapping rule - a relatively simple task, since argument checking amounts to simple LINKed dependency contexts in the CG formalism. The somewhat simplified rule example below targets the verb “tune”:

SUBSTITUTE (V) (<v:for^vtp> <fn:adjust>
 <r:SUBJ:AG> <r:ACC:PAT>)
TARGET ("tune" V)
IF (c @SUBJ LINK 0 <H>) *find daughter dependent (c) subject, check its class*
OR (0 PAS/INF) ... *though this isn't necessary for passives and infinitives*
OR (0 PCP1 + @ICL-N<PRED LINK p <H>) ...
for postnominal gerund clauses, check their mother dependent (p, parent) for human class
AND IF (c @ACC LINK 0 <mach> OR <V>) ...

find accusative daughter (c), check its class
OR (0 PAS LINK c @SUBJ LINK 0 <pass-acc>
 LINK 0 <mach> OR <V>) ... *for passives, check subject class instead*
OR (0 <acc-ellipsis> LINK 1 (*) LINK *-1 @FS-N< BARRIER NON-V ... *in an object-less (<acc-ellipsis>) relative clause (FS-N<*
 LINK p <rel-acc> LINK 0 <mach> OR <V>) ...
find the mother (p) and check its class for machine or vehicle
OR (0 PAS + @ICL-N< LINK p <mach> OR <V>) ... *do the same for postnominal passive clauses*

In this rule, apart from the <fn:adjust> framenet class (implicitly: sense), argument relation tags (<r:....>) are added indicating an AG role (agent) for the subject and a PAT (patient) role for the object, IF the former is human (<H>) and the latter a vehicle (<V>) or machine (<mach>). In the definition section of the grammar, such semantic noun sets are expanded to individual semantic prototype classes (table 3), individual words or a combinations of category tags.

LIST <H> = <H.*>r <hum> <inst> <org> <media>
 <party> <civ> <Lciv> <Ltown> <Lcountry> <Lregion>
 "anybody" "anyone" "everybody" "everyone" "who"
 "one" 1S 2S 2S/P 1P 2P (<fem> PERS) (<mask>
 PERS) (<masc> PERS) ("he" PERS) ("she" PERS)
 ("they" PERS) (<heur> <Proper>);

Table 3: Semantic prototypes

	Semantic (prototype) noun class
<H>	Human: <Hprof>, <Hfam>, <Hnat> <Hideo>
<cc>	concrete object: <cc-stone>, <cc-rag>, <cc-cord> ...
<act>	Action: <act-s> speech-act, <act-do> ... cp. -CONTR: <event> <process>
<L>	Location: <Lh> human place, <Ltop>, <Lwater>, <Labs>, <Lsurf> surface ...
<A>	Animal: <Azo> land animals, <Aorn> birds, <Aich> fish ...
<sem>	Semanticals: <sem-r> book, <sem-l> song, <sem-c> concept, <sem-s> speech ...
<food>	Food: <food>, <food-c>, <food-m>, <fruit> ...
<tool>	Tools: <tool-nus>, <tool-cut> ...
<cm>	Substance: <cm-liq> liquid, <cm-gas>, <cm-chem> ..

<mon>	money
<sit>	situation
<V>	vehicle (<Vground>,<Vair> ...)
<conv>	convention
<HH>	Group: <org>, <media>, <inst> institutions
<an>	anatomical (body part): <anmov>, <anorg>, <anzo>, <anbo> ...
.....	(about 200 classes)

Apart from semantic classes, the frame mapping rules in step one may exploit word class or phrase type (table 4). With noun phrases being the default, special context conditions will be added for finite or non-finite clausal arguments, adverbs and pronouns. Special cases are the 'pl' plural marker (implying np at the same time), and the 'lex' category used for incorporated “as is” tokens.

The second step consisted of the assignment of thematic roles to arguments. Current CG compilers do not allow mappings on multiple (argument) contexts, but with GrammarSoft's open-source CG3 compiler it is possible to unify tag variables with regular-expression string matches, so rules were written to match argument functions with head verb's new <r:....> tags in order to retrieve (and map) the correct thematic role from the latter.

```
MAP KEEPORDER (VSTR:§$1) TARGET @SUBJ
(*p V LINK -1 (*) LINK *1 (<r:.*>r) LINK 0 PAS
LINK 0 (<r:ACC:.(.*)>r);
```

The rule above is a simple example, retrieving a thematic role variable from the verb's accusative argument tag (<r:ACC:....>) and mapping it as a VSTR expression onto the subject in case the verb is in the passive voice. Complete rules will also contain negative contexts (omitted here), for instance ruling out the presence of objects for intransitive valency frames.

The following rule is a generalisation over the @FUNC set (defined in the grammar as objects, predicatives etc. Note that pp roles are mapped on the noun argument of the preposition (@P<) rather than the (semantically “empty”) preposition itself, in spite of the latter being the immediate (syntactic) dependent of the verb. In our CG formalism, such a multi-step dependency relation is expressed as '*p' (open scope parent relation, ancestor relation). The TMP: tags are intermediate tags used for string matches. Thus the additional TMP:§\$2 role tag will be used by rules handling coordination of same-role arguments.

```
MAP KEEPORDER (VSTR:<TMP:§$2> VSTR:§$2)
TARGET @FUNC OR @P< OR @>>P OR <mv>
(0 (<TMP:.*?(\[A-Z]-[+<?>).*?>?>r) LINK *p V
LINK 0 (VSTR:<r:§1:.(.*)>r));
```

While helping to distinguish between verb senses with the same syntactic argument frame, using semantic noun classes as context restrictions raises the issue of circularity in terms of corpus example extraction, and also reduces overall robustness of frame tagging, not least in the presence of metaphor. Therefore, all frame mapping rules are run twice - first with semantic noun class restrictions in place, then - if necessary - without. This way “skeletal-syntactic” (semantics-free) argument structures can still be used as a backup for frame assignment, allowing corpus-based extension of semantic noun class restrictions.

In a vertical, one-word-per-line CG notation, the frame-tagger adds <fn:sense> and <v:valency> tags on verbs, and §ROLE tags on arguments. Free adverbial adjuncts are only partially covered, a few by the frames themselves, but most by separate, frame-independent mapping rules exploiting local grammatical information such as preposition type and noun class. The example demonstrates a frame sense distinction for the English verb *lead*. Dependency arcs are shown as #n->m ID-links.

```
European [European] <*> <jnat> ADJ POS @>N #11-
>12
powers [power] <HH> N P NOM §AG=LEADER
@SUBJ> #12->13
should [shall] <aux> V IMPF @FS-<ACC #13->8
be [be] <vch> <aux> V INF @ICL-AUX< #14->13
leading [lead] <mv> <v:vt> <fn:run_obj>
<fnb:73:Leadership> V PCP1 @ICL-AUX< #15-
>14
the [the] <def> ART S/P @>N #16->18
Western [Western] <jideo> <jgeo> ADJ POS @>N
#17->18
response [response] <event> <act-s> N S NOM
§ACT=ACTION @<ACC #18->15
to [to] PRP @N< #19->18
Russia's [Russia] <*> <Proper> <Lcountry> N S GEN
§AG @>N #20->21
invasion [invasion] <act> N S NOM @P< #21->19
of [of] PRP @N< #22->21
Georgia [Georgia] <*> <Proper> <Lcountry> N S
NOM §PAT @P< #23->22
```

In the example, BFN tags were added to EFN tags, in the form of double role tags, and <fnb:..> frame tags. Independently of the verbal frame lexicon,

the semantic tagger was able to assign an §AG tag to *Russia*, based on the semantic prototype of <act> provided by EngGram with its head noun *invasion*. However, the §PAT tag is a (wrong) default tag – with a true, nominal <fn:invade> frame, it should have been §DES (destination). A future noun frame lexicon should also cover *response*, assigning §CAU (or §STI) to its argument daughter *invasion*.

The second example contains another sense of *lead*, that of *cause*, with §CAU (cause) and §RES (result) as frame arguments. Note the <TRO...> meta tag line providing a time stamp for video alignment. Similar meta mark-up, not shown here, is maintained for speaker, source, topic, news channel etc.

A [a] <*> <indef> ART S @>N #1->3
 blown [blown] ADJ POS @>N #2->3
 tire [tire] <cc-tube> N S NOM §CAU=CAUSE
 @SUBJ> #3->4
 may [may] <aux> V PR @FS-STA #4->0
 have [have] <v.contact> <vtk+ADJ> <aux> V INF
 @ICL-AUX< #5->4
 led [lead] <mv> <v:to^vp> <fn:cause>
 <fnb:5:Causation> V PCP2 AKT @ICL-AUX< #6->5
 to [to] PRP @<PIV #7->6
 <TRO="20080808170708.458">
 this [this] <dem> DET S @>N #8->10
 deadly [deadly] ADJ POS @>N #9->10
 scene [scene] <sem-w> N S NOM §RES=RESULT
 @P< #10->7

Yet another sense of *lead* is that of a path leading somewhere (the *meander*-frame), with §AG and §DES (destination) argument roles. Note that in this third example, the subject agent is not a dependent of *lead* – rather it is the head of the non-finite relative clause in which *lead* is the main verb. We mark such referred roles with an R- prefix (§R-PATH). Also, the first frame in the example illustrates the phenomenon of transparent np's: The direct dependent of *control* is the syntactic object 'all', but semantically this is a transparent (<norole>) modifier part of the argument np, so we raise the semantic function to its 'of X' granddaughter, marking 'roads' as §BEN (beneficiary) of the *run_obj* (*control*) frame. Finally, this is an example of how two roles are necessary on the same token (*roads*), which fill a semantic argument slot in two different frames.

They [they] <*> PERS 3P NOM @SUBJ>
 §AG=CONTROLLING_ENTITY #1->2
 control [control] <mv> <v:vt> <fn:run_obj>

<fnb:1799:Control> V PR -3S @FS-STA #2->0
 all [all] <quant> <norole> INDP S/P @<ACC #3->2
 of [of] PRP @N< #4->3
 the [the] <def> ART S/P @>N #5->6
 roads [road] <Lpath> N P NOM @P< §R-
 PATH=PATH §BEN=DEPENDENT_ENTITY
 #6->4
 leading [lead] <mv> <v:va+DIR> <fn:meander>
 <fnb:61:Path_shape> V PCP1 @ICL-N< #7->8
 into [into] PRP @<SA #8->7
 that [that] <dem> DET S @>N #9->10
 town [town] <Lciv> N S NOM @P< §DES #10->8

N=noun, V=verb, ADV=adverb, INDP=independent
 pronoun, ART=article, DET=determiner,
 KC=coordinating conjunction, PRP=preposition,
 @SUBJ=subject, @ACC=accusative object,
 @ADVL=adverbial, @PIV=prepositional object,
 @SA=subject adverbial, @CO=coordinator, @>N
 prenominal, @N<=postnominal, @FS=finite clause,
 @ICL=non-finite clause, @STA=statement,
 §AG=agent, §PAT=patient, §RES=result,
 §CAU=cause, §DES=destination

4 Evaluation

4.1 Coverage

The reason for using a custom-made Danish-derived FrameNet (EFN) rather than the Berkeley FrameNet (BFN, Baker et al. 1998, Johnson & Fillmore 2000) were not only the better integration of the latter with CG tags and valency frames, but also coverage issues (Palmer & Sporleder 2011). In order to quantify BFN coverage for our speech/news domain, we used an annotated sub-corpus of about 145050 words (of these 19900 punctuation tokens). Due to fall-back strategies, almost all (99.5%) of the 20,343 main verbs in the corpus had been assigned an EFN frame, indicating good basic lexical coverage of the domain. We then checked both the verbs and the frames against BFN v. 1.5. For 26.4% of verb types and 4.1% of verb tokens BFN did not have any frame entry at all². To measure frame coverage, we used BFN frame classes mapped from the assigned EngGram frame categories, checking if the frame in question was associated with a BFN sense for the verb in question. If the verb's valency instantiation matched a valency found in a BFN example sentence, that particular frame had to be one of the EngGram frame classes, making matches more likely. At least with our somewhat heuristic

² Examples were *betray*, *campaign*, *guarantee*, *involve*, *limit* etc.

matching technique, BFN did not have a matching frame in its frame inventory for a given verb in 33.6% of frame instances and for 33.4% of the 1647 frame types in the corpus. This finding supports the analysis by Erk & Padó (2006) that BFN has an unbalanced coverage problem for word senses, with fewer senses per word than the German FrameNet, because it is built one frame at a time, not one verb at a time.

4.2 Performance

To evaluate the coverage and precision of our frame tagger, we annotated a chunk of 882,500 tokens from the UCLA CSA television news corpus, building on an EngGram dependency annotation (Bick 2009) as input, and using only the rules automatically created by our FrameNet conversion program, with no manual rule changes, rule ordering or additions.

Out of 120,843 words tagged as main verbs, 99.9% were assigned a verbal frame sense, though 20.18% of the assigned categories were default senses for the verb in question because of the lack of surface arguments to match for sense-disambiguation. 18.6% of frames were subject-less infinitive and gerund constructions, but of these, 57.2% did have other, non-subject arguments to support frame assignment. The corpus contained 2473 verb lexeme types, and the frame tagger assigned 5840 different frame types, and 4234 verb sense types. Type-wise, this amounts to 2.36 frames, and 1.71 senses per verb type (similar to the distribution in the frame lexicon itself), but token-wise ambiguity is about double that figure, as we will discuss later in this chapter.

Table 4: Frame slot distribution and surface expression probabilities

	frame slots	expressed surface arguments with frame roles	percentage of filled slots
SUBJ	95194	65780 (1061 PCP1 @ICL-N<)	69.1% (da 51.45)
ACC	41765	36629 (978 PAS @ICL-N<)	87.7% (da 77.03)
DAT	1470	1005	68.4% (da 53.72)
PIV	9049	5275	58.3% (da 99.23)
SC	17690	17670	99.9% (da 100.00)

OC	657	446	67.9% (da 100.00)
SA	2809	2762	98.3% (da 100.00)
OA	2231	2021	90.6% (da 100.00)
ADVL	29	27	93.1% (da 100.00)

Table 4 contains a break-down of surface expression percentages for individual argument types. Subject (SUBJ), dative objects (DAT) and prepositional objects appear to be the least obligatory categories, though the latter is lower than it would be in the face of a more unabridged valency lexicon, since the frame mapping grammar also allows pp adverbials to match PIV object slots, to cover cases where the EngGram valency lexicon lacked an entry that the frame mapping grammar did have. It should also be born in mind that both subjects and object slots may be filled not by direct daughters of the main verb, but – for instance – by the heads of non-finite postnominal or finite relative clauses. In these cases, the frame mapping grammar may encounter a slot filler without leaving a mark on the syntactic function counter used for to compute the above table. Predicative arguments (SC), of verbs like *be* and *become*, are 100% expressed, as are valency-bound adverbials (SA). and prepositional arguments (PIV) have almost as high an expression rate simply because most verbs have alternative valency frames of lower order (intransitive or monotransitive accusative) that the tagger would have chosen in the absence of a PIV argument. In other words, PIV arguments are strong sense markers, and their absence will sooner lead to false-positive senses of lower valency-order than to PIV-senses without surface PIV. Among the safest markers for frame senses are incorporated particles (§INC), as in *give up*, *take place etc.*, which are almost 100% obligatory for the valency pattern in question, and which the frame mapping grammar therefore will try to match these before more general verbal complementations.

On a random 5000-word chunk of the frame-annotated data, a complete error count was performed for all verbs. All in all, there were 629 main verb tags, of which 13 should have been auxiliaries and one had been wrongly verb-tagged by the parser (*even*). Our frame tagger assigned 624 frames, missing out on only 4 regular verbs (2 *x vetted*, *unquote*, *harkens*), and (wrongly) tagging the false-positive verb. This suggests a very good coverage in simple lexical terms (99.6%). In 20.7% of cases, the frame tagger assigned a default frame,

usually a low-order valency frame without incorporates³. Of 615 possible frames, 495 were correctly tagged, yielding the following correctness figures:

Table 5: Performance of the frame sense tagger on television news

	Recall	Precision	F-score
total	80,49 (da 85.05)	79.32% (da 85.20)	79.91
ignoring pos/aux errors	80.49%	81.01%	80,75

These figures are an encouraging result, despite the “weak” (inspection-based) evaluation method. The performance falls 5 percentage points short of the results achieved for our point of departure, the Danish FrameNet (Bick 2011), but it has to be borne in mind that the current English system is work in progress, as indicated in rough terms by the smaller lexicon size of the new, derived *framenet*. More important than lexicon size as such, is granularity – the coverage of frequent verbs in term of valency frames and incorporations – and here the method of trying to port frames across languages was bound to miss out on many English constructions simply because translations tend to be many-to-one, i.e. conflating several rarer SL constructions to one or few more general TL constructions. Using ML techniques, the best participating systems in the SemEval 2007 frame identification task (Baker, Ellsworth & Erk 2007) achieved F-scores between 60% and 75%, though because of the stricter evaluation system any direct comparison will have to wait for future work based on a category mapping scheme, using the same data. Shi & Mihalcea (2004), also using FrameNet-derived rules, report an F-score of 74.5% for English, while Gildea & Jurafsky (2002), using

³ The default frame is not currently based on statistics, but decided upon when converting the *framenet* lexicon into a Constraint Grammar, as the first intransitive or monotransitive valency frame by order of appearance in the lexicon,. Other valency categories may also be default-rated, but need a special manual tag (atop=”1”). Similarly, frames can be downgraded by assigning them higher atop numbers – in effect meaning they will only be used if all context slot conditions are present in the sentence. This way, the ultimate ranking of frames, and the decision on a default frame is fully controlled by the lexicographer.

statistical methods, report F-scores of 80.4% and 82.1% for frame roles and abstract thematic roles, respectively. For copula and support verb constructions, not included in the earlier evaluations, Johansson & Nugues (2006) report tagging accuracies for English of 71-73%, respectively, but a comparison is hard to make, since we only looked at support constructions that our FrameNet does know, with no idea about the theoretical lexical “coverage ceiling”.

A qualitative look at the errors shows that the underlying part-of-speech tagging was very robust – thus only 2 verb class errors were found, one false positive and one false negative. The confusion of auxiliary and main verb for *be* and *have*, however, did play a certain role (10% of false positive frames), and so did incomplete valency frames or wrong syntactic attachments, resulting in missing slot fillers for the frame-mapping rules. Some of these underlying errors were ultimately domain-dependent and due to non-standard language in our (spoken) corpus. Thus, half of the auxiliary/main verb-confusion occurred due to missing words (*have bombing* instead of *have been bombing*) or unfinished sentences or retractions (*I don't - I think my people ...*). Ignoring these errors, i.e. assuming correct tagging input, would influence precision, in particular, and raise the overall F-score by 1-2 percentage points⁴.

A break-down of error types revealed that about 40% of all false positive errors (but only 8% of all frames) were cases where the human “gold sense” was not (yet) on the list of possible senses in the our EFN database. As one might expect, default mappings accounted for a higher percentage (26.4%) among error verbs than in the chunk as a whole (20.2%), and contributed to almost a third of the “frame-not-in-lexicon” cases.

Frequent verbs have a high sense ambiguity, and verbs with a high sense ambiguity were more error-prone than one-sense verbs, as can be seen from the table below. Thus, the verbs occurring in our evaluation chunk had 4.7 potential senses per verb (6.54 for the ambiguous ones), and the verbs accounting for frame tagging errors had a theoretical 10.26 senses each. While these numbers and their proportions closely matched the findings for Danish, there is a marked difference in the “sense density” for the verb lexica as such, reflecting the fact that the larger size of the Danish

⁴ Given the syntactic and semantic knowledge base of our system, it would be feasible to design rules for identifying “false” main verbs at a later stage, to remedy this problem.

FrameNet in terms of verb types is achieved by including the Zipf tail of verbs – i.e. rare verbs with one or few readings – while the overall sense count is not so different. Concluding from this, one can assume that an enlargement of EFN in terms of verb types will decrease rather than increase the ambiguity strain on tagging performance.

Table 6: Sense ambiguity per verb

	Verb type count	theoretical sense count	senses / verb	sense type count in chunk (as tagged)
EFN frameNet lexicon	4774	10800	2.26 (da 1.46)	-
verb types in chunk	205	964	4.7 (da 4.21)	244
sense ambiguous	140	916	6.54 (da 6.77)	193
frame error verbs	56	576	10.26 (da 10.08)	78

5 Conclusions and Outlook

We have shown that a robust semantic tagger for English television news can be built by converting a valency-anchored fragment into Constraint Grammar mapping rules, turning syntactic and semantic selection restrictions into dependency-linked context conditions. Though the system has a reasonable lexical coverage and frame sense recall for verbs, a great deal of work needs to be done on nominal frames and verbo-nominal incorporations. Also, evaluation should be carried out for semantic role tagging accuracy in addition to verb senses, optimally in a standardized evaluation environment.

References

- Bick, Eckhard. 2000. *The Parsing System Palavras - Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework*, Aarhus: Aarhus University Press
- Bick, Eckhard. 2009. "Introducing Probabilistic Information in Constraint Grammar Parsing". *Proceedings of Corpus Linguistics 2009, Liverpool* (ucrel.lancs.ac.uk/publications/cl2009/)
- Bick, Eckhard. 2007. Automatic Semantic Role Annotation for Portuguese. In: *Proceedings of TIL 2007* (Rio de Janeiro, July 5-6, 2007). ISBN 85-7669-116-7, pp. 1713-1716
- Bick, E. & H. Mello & A. Panunzi & T. Raso (2012), The Annotation of the C-ORAL-Brasil through the Implementation of the Palavras Parser. In: Calzolari, Nicoletta et al. (eds.), *Proceedings LREC2012* (Istanbul, May 23-25). pp. 3382-3386. ISBN 978-2-9517408-7-7
- Baker, C., Ellsworth, M., & Erk, K. 2007. SemEval-2007 Task 19: Frame Semantic Structure Extraction. *Proceedings of (SemEval-2007), ACL 2007*, pages 99-104
- Baker, Collin F., Fillmore, J. Charles & John B. Lowe. 1998. The Berkeley FrameNet project. In *Proceedings of the COLING-ACL, Montreal, Canada*
- DeLiema, David & Francis Steen & Mark Turner. 2012. *Language, Gesture and Audiovisual Communication: A Massive Online Database for Researching Multimodal Constructions*. Lecture. 11th Conceptual Structure, Discourse and Language Conf., Vancouver, May 17-20.
- Erk, Katrin & Sebastian Padó. 2006. SHALMANESER – A Toolchain for Shallow Semantic Parsing. *Proceedings of LREC 2006*
- Gildea, D. and D. Jurafsky. 2002. Automatic Labeling of Semantic Roles, *Computational Linguistics*, 28(3) 245-288.
- Johansson, Richard & Pierre Nugues. 2006. Automatic Annotation for All Semantic Layers in FrameNet. *Proceedings of EACL 2006*. Trento, Italy.
- Johnson, Christopher R. & Charles J. Fillmore. 2000. The FrameNet tagset for frame-semantic and syntactic coding of predicate-argument structure. In *Proceedings of ANLP-NAACL 2000*, April 29-May 4, 2000, Seattle WA, pp. 56-62.
- Karlsson, Fred, Atro Voutilainen, Juha Heikkilä, and Arto Anttila. 1995. *Constraint Grammar: A Language-Independent System for Parsing Unrestricted Text*. Natural Language Processing, No

4. Mouton de Gruyter, Berlin and New York

- Kipper, Karin & Anna Korhonen, Neville Ryant, and Martha Palmer. 2006. Extensive Classifications of English verbs. *Proceedings of the 12th EURALEX*. Turin, Sept. 2006.
- Levin, Beth. 1993. *English Verb Classes and Alternation, A Preliminary Investigation*. The University of Chicago Press.
- Palmer, Alexis & Caroline Sporleder. 2011. Evaluating FrameNet-style semantic parsing: the role of coverage gaps in FrameNet. *Proceedings of COLING '10*. Pp 928-936. ACL.
- Palmer, Martha, Dan Gildea, Paul Kingsbury. 2005. The Proposition Bank: An Annotated Corpus of Semantic Roles. *Computational Linguistics*, 31:1., pp. 71-105, March, 2005.
- Shi, Lei & Rada Mihalcea. 2004. Open Text Semantic Parsing Using FrameNet and WordNet. In HLT-NAACL 2004, Demonstration Papers. pp. 19-22