An empirically grounded expansion of the supersense inventory

Héctor Martínez Alonso† **Anders Johannsen** **Sanni Nimb**‡
Sussi Olsen **Bolette Sandford Pedersen**†
University of Copenhagen (Denmark)
‡Danish Society of Language and Literature (Denmark)
†alonso@hum.ku.dk, bspedersen@hum.ku.dk

Abstract

In this article we present an expansion of the supersense inventory. All new supersenses are extensions of members of the current inventory, which we postulate by identifying semantically coherent groups of synsets. We cover the expansion of the already-established supernsense inventory for nouns and verbs, the addition of coarse supersenses for adjectives in absence of a canonical supersense inventory, and supersenses for verbal satellites. We evaluate the viability of the new senses examining the annotation agreement, frequency and co-ocurrence patterns.

1 Introduction

Coarse word-sense disambiguation is a well established discipline (Segond et al., 1997; Peters et al., 1998; Lapata and Brew, 2004; Alvez et al., 2008; Izquierdo et al., 2009) that has acquired more momentum in the latter years under the name of *supersense tagging* (SST). SST uses a coarse sense inventory to label spans of variable word length (Ciaramita and Johnson, 2003; Ciaramita and Altun, 2006; Johannsen et al., 2014). This coarse sense inventory is obtained from the list of WordNet *first beginners*, i.e. the names of the *lexicographer files* that hold the synsets.

However, lexicographer files were devised for practical reasons, namely as an organization method for the development of WordNet (Miller, 1990; Gross and Miller, 1990; Fellbaum, 1990), and not as final target categories to annotate with or disambiguate from.

Nevertheless, the organization of lexicographer files is semantically motivated, and supersenses have proven useful for natural language processing such as metaphor detection or relation extraction (Ciaramita and Johnson, 2003; Tsvetkov et al., 2014a; Søgaard et al., 2015). According to Ciaramita and Altun (2006), supersenses extend the named entity recognition (NER) inventory so that the predictions of an SST model subsume the output of NER. Schneider et al. (2015) provide a full SSI for prepositions.

The current supersense inventory (henceforth SSI) enjoys *de facto* standardness, but in spite of its potential usefulness, it is used acritically. The current SSI provides 26 noun supersenes and 15 verb supersenses. Adjective and adverb lexicographer files are disregarded. We provide a revision of the SSI by an extension of its supersenses using the Danish wordnet as starting point.

This revision is empirically backed by four evaluation criteria, namely inter-annotator agreement, sense frequency after adjucation, sense co-ocurrence, and NER compliance (whenever possible). Note that we do not suggest merging existing supersenses, but only extending the current SSI in a backwards-compatible manner.

We conduct our extension in three steps steps. First, we propose new supersenses when a projection between an EuroWordNet (EWN) ontological type and a supersense is not univocal (Section 2). Second, we evaluate the distribution of supersenses in terms of agreement after an annotation task, frequency and sense-sense relations (Section 4) and analyze the results across the different parts of speech (Section 5). Lastly, we suggest new supersenses (underlined in in Table 2) when large sections of the data have been assigned to backoff categories.

The main contributions of this paper are i)a set of guidelines for the inclusion of new supernses in the SSI, ii) an empirically motivated expansion of the SSI with new senses for nouns, verbs and adjectives respectively,[1] and iii) a projection from ontological types to supersenses that can be used to enrich any wordnet that is not organized in lexi-

[1] https://github.com/coastalcph/semdax

cographer files or where synsets are not fully connected to Princeton synsets.

2 Extending the supersense inventory

This section describes the extension of the SSI that results from an analysis of projections into supersenses from ontological types, ensuing both retro-compatibility with the existing inventory (i.e. all new supersenses are extensions of an existing supersense), and compability with NER tags.

We use The Danish wordnet (Pedersen et al., 2009), DanNet, as a starting point. DanNet is not organized in lexicographer files. However, its synsets are associated to ontological types (Vossen et al., 1998). We map from the ontological type of the synsets to a supersense. Table 2 provides one example for each lexical part of speech.

Ontological type	Supersense
Property+Physical+Colour	ADJ.PHYSICAL
Liquid+Natural	NOUN.SUBSTANCE
Dynamic+Agentive+Mental	VERB.COGNITION

Table 1: Supersense mapping examples.

We establish a projection into supersenses with the following steps; if an ontological type t_i:

1. does not have a straightforward 1-to-1 mapping to a supersense,
2. is the subtype of an ontological type t_j (e.g. *Liquid+Natural* is a subtype of *Liquid*),
3. and has enough support (in terms of how many synsets make up t_i),

then we propose new supersense for t_i as an extension of the supersense of t_j. We consider the support to be substantial enough when a subtype has at least 500 synsets out of the 65k synsets in DanNet and, and it makes up at least 12% of its parent supersense.

We exemplify this method by explaining how we extend DISEASE from STATE. The subtype *Property+Physical+Condition* is associated to 527 synsets and makes up 70% of the synsets of the type *Condition*. All the synsets of this subtype are diseases, and we propose the supersense DISEASE as an extension of STATE, which is otherwise the supersense translation of *Condition*.

In addition to providing new supersenses for the main three lexical parts of speech, we devise three aditional tags for verbal satellites (collocations, particles and reflexive pronouns) as aid for verbal

New supersense	Subsumed by
Noun	
VEHICLE BUILDING CONTAINER	} ARTIFACT
DOMAIN ABSTRACT	} COGNITION
INSTITUTION	} GROUP
DISEASE	} STATE
LANGUAGE DOCUMENT	} COMMUNICATION
Verb	
ASPECTUAL	} STATIVE
PHENOMENON	} CHANGE
Adjective	
MENTAL PHYSICAL SOCIAL TIME FUNCTION	} ALL
Satellite	
COLLOCATION PARTICLE REFLPRON	} *none*

Table 2: Extensions to the sense inventory. Items in grey do not fulfill the inclusion criteria, underlined items have been suggested during post-annotation analysis.

multiwords the annotation (cf. Section 5.4). Table 2 lists the new supersenses. Underlined dupersenses marked are determined in post-annotation analysis (cf. Section 5), while the rest have been determined during the projection step described in this section. Supersenses in grey do not meet the inclusion criteria, and are thus not incorporated in our proposal for SSI extension.

3 Annotation task

We perform an annotation task on 5,500 sentences from a Danish contemporary corpus (Asmussen and Halskov, 2012) made up of newswire, parliamentary speech, blog posts, internet forum discussions, chatroom logs and magazine articles, plus the test section of the Danish Dependency Treebank (Buch-Kromann et al., 2003).

Any corpus choice imposes a bias, and we base the corpus choice on a twofold need: to tune the sense inventory to the needs of contemporary genres that are used for information extraction, without sacrificing its adequacy for more usual domains. Generally speaking, another corpus choice would yield a different supersense expansion.

The corpus was pre-annotated using the supersense projection list described in Section 2. Even though the size of the specific wordnet is a determining factor for the quality of the preannotation, it does not determine the coverage of the final supersense annotation, which provides full coverage because a SSI covers all content words.

Two in-house native annotators with a background in linguistics annotated the data, choosing the best pre-annotated sense or selecting a new one. A third annotator performed adjudication in case of disagreement. The overall kappa score before adjudication is 0.62. Olsen et al. (2015) provide more details on the annotation task. The resulting data has been use for automatic supersense tagging by Martínez Alonso et al. (2015).

4 Metrics

This section describes the metrics applied to the supersense-annotated corpus in order to assess the distribution of the new supersenses.

4.1 Sense-wise agreement variation

Inter-annotator agreement is a source of information on the reliability of semantic categories (Lopez de Lacalle and Agirre, 2015). In this section, we examine the variation in agreement for noun and verb supersenses. Cf. Olsen et al. (2015) for a more detailed account.

Figures 1 and 2 portray the variation of agreement across noun and verb supersenses. Each cell in the matrix indicates the probability of a token being annotated with a row-column tuple of supersenses (r_i, c_j) by the two annotators. The matrix is normalized row-wise, and each row describes the probability distribution of a certain supersense r_i to be annotated with any other supersense c_j. When r_i and c_j have the same value, annotators agree. Rows are sorted in descending order of agreement, i.e. the size of the $r_i = c_j$ box on the diagonal. The larger the box in the diagonal, the higher the agreement for a given r_i supersense.

From the standard supersenes, for instance, N.GROUP is very seldom assigned by both annotators, and there is usual disagreement with N.QUANTITY. Other senses like N.BODY have very few off-diagonal values and have near-perfect agreement.

Out of the new supersenses, N.INSTITUTION has very high agreement. However, the new supersense N.DOMAIN has very low agreement. A domain (i.e. a field of knowledge or professional discipline) is difficult to distinguish from its semantically related senses N.COGNITION and N.COMMUNICATION. Low agreement also compromises the reliability of some of the established supersenses such as NOUN.SHAPE. However, the goal of these measurements is to evaluate the new supersenses, because we do not advocate for a reduction of the canonical SSI, but an extension of the existing list of supersenses.

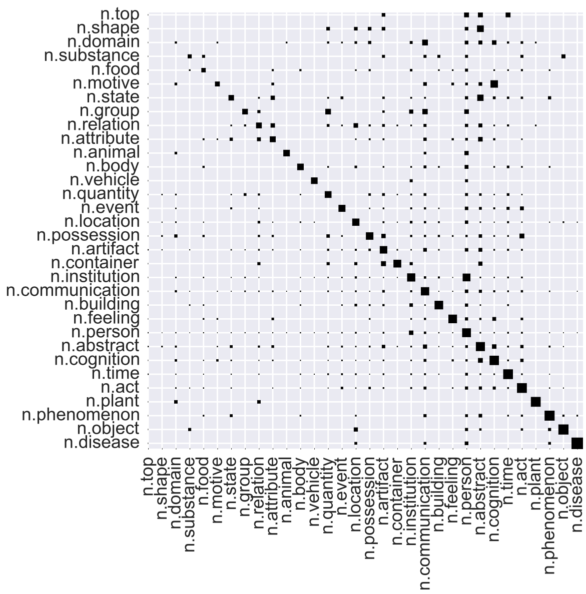

Figure 1: Agreement variation for nouns.

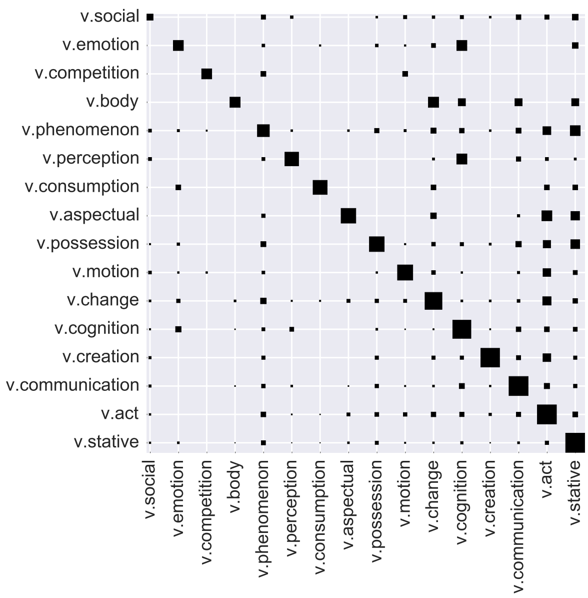

Figure 2: Agreement variation for verbs.

Agreement also varies across parts of speech. Diagonal boxes take up 69% of the probability mass of the verbs, while 58% is taken by the agreed nouns. In other words, 31% of the annotations for verbs are mismatched, whereas 42% of the nouns have mismatching annotations. We consider this difference a consequence of the size of the inventory for nouns and verbs respectively, and not an indication of verbs being *per se* easier to annotate than nouns.

4.2 Supersense frequency

Frequency is the most straightforward way of assessing whether a a certain sense has been given to enough examples to be considered relevant. If a new sense is very frequent, there is sufficient reason to consider it as a valid addition to the SSI.

Table 3 provides the absolute frequency for the 28 most frequent supersenses, namely half of the total SSI, after disagreements had been resolved by the adjudicator.

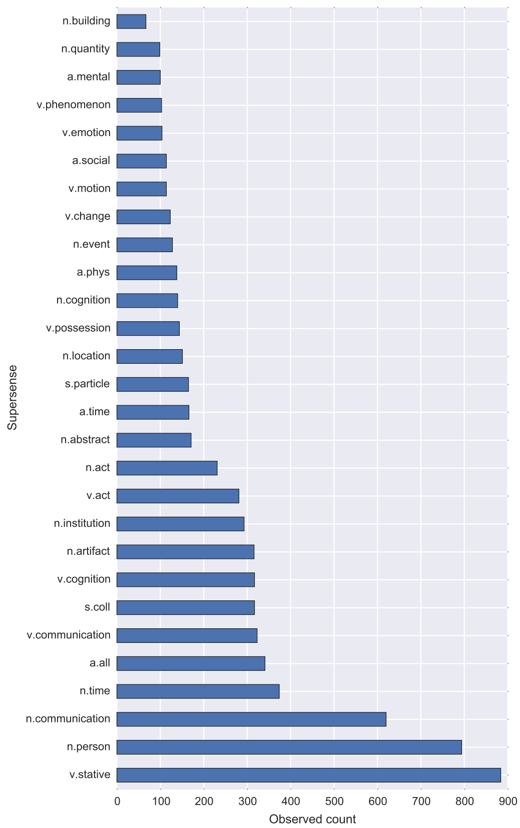

Figure 3: Distribution of frequent senses.

Presence in the top half of the sense ranking is one of the criteria for inclusion in the SSI.

4.3 Association between supersenses

A third source of information on the appropiateness of a supersense is its relation with the other established senses. This section offers an overview on how supersenses co-occur. To give account for relevant associations between senses, we use PMI (pointwise mutual information). Higher PMI values indicate stronger association, i.e. a higher conditional probability of one sense appearing in a sentence given the other, controlled for the frequency of both senses in order not to overestimate the co-occurrence of frequent senses.

Table 3 shows the twelve pairs of supersense with the highest PMI calculated across sentences. We compare the supersense-wise PMI for three corpora:

1. Danish extended (DA-EX): The Danish corpus annotated with the extended SSI described in Section 3,
2. Danish regular (DA-RG): The Danish corpus from Section 3 with regular supersenses, where the extended senses have been replaced by their subsuming original sense, e.g. all the occurrences of N.VEHICLE in DA-EX are N.ARTIFACT in DA-RG,
3. English regular (EN-RG): The English SemCor (Miller, 1990) with the regular supersense annotation.

Some of the associations are prototypical selectional restrictions like V.COMSUMPTION + N.FOOD. Other associations are topical across parts of speech, like VERB.COMPETITION and NOUN.EVENT ('They **won** the **final**'). Finally, there are associations within a part of speech, like N.DISEASE and N.BODY, or N.FOOD and N.CONTAINER. In these associations, one sense is a strong indicator for the other at the topic level (diseases are bodily, food is kept somewhere, etc).

In DA-EX we observe that three of the new nominal senses appear strongly associated with standard supersenses. These relations are topical and easy to interpret. The vehicle-substance relation is the least straightforward one and describes vehicles and the fuel they use, or the materials they are built from.

Projecting back to the regular SSI is not equivalent to annotating from scratch with it. Nevertheless, if we examine the top supersense pairs for DA-RG, we observe that the V.STATIVE sense appears three times. By ignoring the aspectual differ-

Danish (extended)		Danish (regular)		English (regular)	
v.consumption	n.food	v.consumption	n.food	v.consumption	n.food
v.contact	n.body	**v.stative**	n.plant	v.weather	n.object
n.food	**n.container**†	n.person	n.animal	v.weather	n.phenomenon
v.body	n.body	v.competition	n.relation	n.plant	n.food
n.disease†	n.body	v.competition	v.event	n.plant	n.animal
v.competition	n.event	v.change	n.substance	n.substance	n.process
v.motion	v.contact	n.state	n.feeling	v.body	n.body
v.contact	n.artifact	v.consumption	v.change	v.weather	n.substance
n.substance	n.object	v.motion	n.object	v.emotion	n.motive
n.shape	n.body	**v.stative**	v.consumption	n.plant	**n.tops**
n.vehicle†	n.substance	**v.stative**	n.substance	v.contact	n.body
v.competition	n.relation	n.substance	n.person	n.food	n.animal

Table 3: Sense pairs ranked by PMI, bold and underlined described in Section 4.3, † marks new sense.

ence, the tag receives associations with N.PLANT, V.CONSUMPTION and N.SUBSTANCE. Upon manual examination we deem these relations to be spurious, i.e. caused by the presence of the verb *være* ('be') somewhere in the sentence, except the relation between V.STATIVE and V.CONSUMPTION, which is aspectual in nature. The effect on the distribution of supersenses when projecting back to the original SSI becomes apparent for the pair V.COMPETITION + N.RELATION, which becomes the fourth highest PMI in DA-RG.

The English supersense associations of EN-RG provide an example on the effect of corpus choice when annotating. The fairly uncommon N.PLANT appears in several of the top associations, which is a sign of plant senses being used in very restricted contexts in this corpus (biology and recipes). Moreover, we also find a strong association with one of the backoff senses, namely N.TOPS, which is not desirable.

5 Supersenses across parts of speech

5.1 Nouns

This section describes the extended SSI for nouns. To the extent that nouns denote entities, they are very often of focus of interest of ontologies. To the extent that entities often have physical denotation—and thus concrete meaning—, they are the easiest concepts to categorize semantically. Indeed, many ontologies are largely nominal, cf. Suchanek et al. (2008) or Wu and Weld (2008).

WordNet lexicographer files were developed before the consolidation of NER, and named-entity coverage in wordnets is irregular. If, as stated in Section 1, NER compatibility is a favorable side effect of SST, we consider improved NER compatiblity of the new SSI as a plus.

Even though NER inventories are application dependent (cf. Nadeau and Sekine (2007) for a survey), our reference is the *de facto* standard CONLL inventory (Tjong Kim Sang and De Meulder, 2003), with the labels PERSON, LOCATION and ORGANIZATION, as well as a MISCELLANEOUS label, needed for full coverage but not present in e.g. the 7-label inventory of MUC-7 (Chinchor and Robinson, 1997).

Concrete meaning is easier to annotate (Passonneau et al., 2009) and can be the easiest to extend with new senses. As a matter of fact, the concrete N.ARTIFACT supersense is the one that yields more new supersenses in our analysis, namely N.BUILDING, N.CONTAINER and N.VEHICLE. In particular, N.BUILDING extends N.ARTIFACT because artifactual locations, already noted as a semantic type the SIMPLE ontology (Lenci et al., 2000), like houses and highways are very often predicated as locations (following locative prepositions, etc.) instead of having the typical distribution of artifacts, i.e. with the verb *use* or the preposition *with*. Moreover, N.BUILDING maps better into the *Location* type of NER. We leave the potential supersenses for instruments and machines as parts of N.ARTIFACT and do not specifiy them even further, because they hold the prototypical meaning of the supersense.

In spite of the expected higher difficulty of dealing with abstract meaning, we examine two extensions for the abstract supersense N.COGNITION yielded by the the ontological type projection from Section 2, namely N.DOMAIN and N.ABSTRACT. The supersense N.DOMAIN covers fields of knowledge such as *philosophy*, but also other disciplines to cover sense alternations like 'I enjoyed this **dance**' (N.ACT) vs. 'I studied **dance** at the Performing Arts Academy' (N.DOMAIN). The supersense N.ABSTRACT aims at covering concepts like *idea*, and as a label for metaphorical usages of other concrete words like *pattern* in 'behavioral pattern'.

The fairly abstract supersense N.STATE yields a concrete sense DISEASE, which is much easier to annotate than its original parent supersense (cf. Figure 1). Lastly, we extend N.GROUP with N.INSTITUTION. The original sense does not map neatly into NER, as the overlap is only partial; while *ministry* would fall under the ORGANIZATION type of NER, *pack* (of rats) and *school* (of fish) would not.

5.1.1 Sense-wise evaluation

In this section we evaluate the extended noun supersenses according to four properties summarized in Table 4; whether the agreement for a supersense is high enough (Agr.), whether its frequency is high enough, whether we identify relevant associations using PMI (Assc.), and whether it potentially improves NER compliance (NER). Moreover, we suggest two new supersenses, N.LANGUAGE and N.DOCUMENT, indicated in the lower section of Table 4.

The first three properties are obtained from the metrics in Section 4. We consider agreement to be high enough when there is at least 51% agreement for a supersense. We consider frequency to be enough when the sense belongs to the first 28 senses out of 56 (i.e. the first half of the frequency-ranked SSI). None of the thresholds are particularly high, but we consider a noun supersense as a candidate for inclusion in the final SSI if two of the four properties are satisfied. In other words, none of the criteria are necessary, but fulfilling two of them is sufficient.

We observe most of the new senses fulfill at least two of the criteria, with the exception of N.DOMAIN, which fulfills none. Thus, we do not endorse using the N.DOMAIN supersense and still use N.COGNITION for fields of knowledge. Nevertheless, the N.ABSTRACT sense seems a valuable extension because it satisfies the agreement and frequency criterion.

New supersense	Agr.	Freq.	Assc.	NER
ABSTRACT	x	x		
BUILDING	x			x
CONTAINER	x		x	
DISEASE	x			x
DOMAIN				
INSTITUTION	x	x		x
VEHICLE	x		x	
LANGUAGE	–			
DOCUMENT	–	x		x

Table 4: Inclusion criteria for new noun senses.

The strongest nominal candidate for inclusion is N.INSTITUTION, which satisfies the first two initial criteria, plus improves NER compatibility.

During the annotation task, we observed that a large amount of examples of the standard N.COMMUNICATION supersense were document names, movie titles, and so on. One of the authors of this article reviewed all the N.COMMUNICATION spans and classified them in three categories, two of them mapped from the EWN top ontology, N.DOCUMENT and N.LANGUAGE, and a third back-off category for N.COMMUNICATION. Notice how, in spite of having spawned three senses (N.CONTAINER, N.VEHICLE and N.BUILDING), N.ARTIFACT is still a very frequent supersense.

The document-language distinction is a high-level type in the SIMPLE ontology (Lenci et al., 2000). Note that these two new communication subsenses do not solve the artifact-information ambiguity commonly found in lexical semantics (Pustejovsky, 1991). While N.LANGUAGE has more often an eventual reading (e.g. *conversation*, *remark*), N.DOCUMENT refers more often to works and other entities with a non-temporal denotation. We also use N.LANGUAGE for the metalinguistic usage of words (e.g. 'The word **drizzle** sounds funny'). This re-annotation produces examples like the following:

> H. C. Andersen er jo verdensberømt , fordi hans **forfatterskab**/N.DOCUMENT er blevet oversat til alle sprog/N.LANGUAGE .
>
> *H. C. Andersen is world famous, because his* ***writing*** *has been translated to all languages.*

Out of the 1513 N.COMMUNICATION cases, 360 fall under N.LANGUAGE and 928 under

N.DOCUMENT, and the remaining were left with the original label. Out of the 929 N.DOCUMENT spans, 382 are named entities, where 248 are +2 tokens in length. This metric aims at justifying having document as an NER label, where span identification is as relevant as proper labeling.

We believe the frequency of document-name named entities makes a good case for considering the N.DOCUMENT class as an addition to the SSI and to NER. However, we do not find enough support to recommend a N.LANGUAGE supersense and prefer using the original N.COMMUNICATION instead.

5.2 Verbs

Verbs are central to the theory of lexical semantics, yet their semantic characterization has been closer to the syntax-semantics interface (Levin, 1993; Kipper et al., 2000; Kipper et al., 2006). In this aspect, the wordnet SSI for verbs is very different, e.g. verbs like *jump* or *displace* are of the V.MOTION, even though their argument structures are very different. Nevertheless, verbal sense alternations are often associated with different argument structures (Grimshaw, 1990).

The V.CHANGE supersense is populated with semantically disparate categories and is very difficult to annotate, even though it is a very frequent sense, both in terms of annotated words and of synsets adscribed to it. According to Fellbaum (1990), 'the concept of change is flexible enough to accomodate verbs whose semantic description mathen them unfit for any other semantically coherent group'. In other words, the rummage box category for verbs is actually the majority class. Indeed, an expansion of change into its subsenses of CHANGE-VARY, CHANGE-STATE, CHANGE-REVERSAL, CHANGE-INTEGRITY, CHANGE-SHAPE and CHANGE-ADAPT could potentially make the supersense more useful, if one is willing to incur the cost of annotating with five more labels.

The V.PHENOMENON supersense extends V.CHANGE by delimiting events that have no agency and are not weather-related, such as *happen*, or *occur*. WordNet shows a systematic ambiguity between V.STATIVE and V.CHANGE for aspectual readings of verbs, and we also propose V.ASPECTUAL for constructions like '**start** the engine' or '**begin** to hope'.

We evaluate verb sense using the criteria we used for nouns in Section 5.1, but discarding NER compliance, which does not apply to verbs. Table 5 shows the criteria for verbs.

New supersense	Agr.	Freq.	Assc.
ASPECTUAL	x		x
PHENOMENON	x	x	

Table 5: Inclusion criteria for new verb senses.

Both new verbal supersenses satisfy two out of three of the criteria, and we can consider them candidates for the SSI extension. We leave it for further discussion whether aspectual verb reading deserves a full-fledged supersense or should be used as a satellite tag (cf. Section 5.4).

5.3 Adjectives

SST as defined by Ciaramita and Johnson (2003) only labels nouns and verbs. Adjectives have received much less attention than nouns and verbs, arguably because of the inherent difficulty of their analysis, cf. Boleda et al. (2012) for a survey on adjective classifications. In addition to the theoretical complications, adjectives are not regarded as core elements of meaning when building applications. For instance, in WordNet 3.0 there are 82k synsets for nouns, 14k for verbs, 18k for adjectives and 4k for adverbs. However, the base concepts from EWN (Vossen et al., 1998), with 4,869 synsets in total, hold 37 adjectives in contrast to 3,210 nouns and 1,442 verbs.

Moreover, the supersense-synset relation is hyponimic, but adjectives in WordNet are not taxonomically organized (Gross and Miller, 1990). For instance, there is no way to retrieve that *ashamed* and *exasperated* are emotional in nature (Tsvetkov et al., 2014b).

The meaning plasticity of adjectives makes it also hard to determine whether adjectives hold any meaning onto themselves, or their meaning is an emergent property of the relation they establish with the noun they complement. Murphy and Andrew (1993) consider adjectives monosemous elements that define their sense when predicated alongside nouns. Under this light, supersense adjectives would be superflous if adjective meaning is an epiphenomenon of noun meaning.

However, insofar adjectives can help disambiguate nominal polysemy (Tsvetkov et al., 2014a), and have different listed synsets, we advocate for providing a set of supersenses for adjectives. This addition makes therefore SST truly all-words for the three main lexical parts of speech. Adjective classifications into supersenses or coarse classes do exist, notably in GermaNet (Hamp and Feldweg, 1997), which Tsvetkov et al. (2014b) apply to English.

When applying the projection method from Section 2, we extend A.ALL with A.MENTAL, A.PHYS, A.SOCIAL and A.TIME. These supersenses do not distinguish descriptive (i.e. extensional) from reference-modifying (intensional) adjectives, e.g. *former* is A.TIME while *imaginary* is A.MENTAL. These senses do not distinguish relational adjectives either, to the extent that *ecologic* and one of the senses of *green* should fall under the same supersense.

The new adjective SSI cannot be evaluated in the samme manner as nouns. The adjective SSI is much smaller, and the agreement and frequency metrics can be misleadingly positive. Indeed, all adjective supersenses satisfy the agreement and frequency criteria specified in Section 5.1.1.

However, A.ALL is the most frequent supersense for adjectives, and it covers 40% of the annotated adjectives. This proportion is too large, and indicates the sense inventory needs to be further specified in order to minimize how many tokens get assigned the backoff sense.

Many of the adjectives under A.ALL are function-appraisal related, such as *god* ('good'), *bedre* (better'), *stor* ('large' as in 'grand'), *vigtig* ('important'). While polarity is an important property of adjectives (Chesley et al., 2006), we do not consider it a desirable trait for supersenses, which are more oriented towards conveying sense denotation that connotation. Hence, we suggest a new supersense A.FUNCTION to give account for function-related senses, what in the terminology of Pustejovsky (1991) would be the *telic role*. We observe that the ALLGEMEIN ('general') category of GermaNet and Tsvetkov et al's MISCELLANEOUS hold similar senses.

5.4 Satellites

When annotating nouns in Section 3, we annotate continuous NER-like spans. But verb-headed multiwords pose a challenge because they are not necessarily continuous, and pose attested challenges for their annotation and automatic recognition (Hoppermann and Hinrichs, 2014; Baldwin, 2005b; Baldwin, 2005a).

We use three satellite tags; S.COLLOCATION, S.PARTICLE and S.REFLPRON (for reflexive pronouns). While the particle distinction is more relevant for satellite-framed languages (Talmy, 1985) like Germanic languages, light-verb constructions are pervasive in many languages, also characteristically verb-framed languages like Spanish or French, where we find verb-headed multiwords like *llevar a cabo* (lit. 'take to ending', 'carry out') or *avoir l'air* (lit. 'to have the air', 'seem'), respectively. A similar approach has been used by Schneider and Smith (2015).

The intention of these tags is to help isolate the head of a verb-headed multiword. We assign the sense label to the syntactic head, even though a light verb construction would be arguably best headed by its introduced noun. In this manner, *gøre grin af* ('make fun of') would be labeled as *gøre*/V.COMMUNICATION *grin*/S.COLLOCATION *af*/S.COLLOCATION', and we thus avoid giving *gøre* ('make') the V.CREATION sense.

6 Conclusions and further work

We suggest an extension of the SSI for the three main lexical parts of speech. We obtain new supersenses using a mapping from ontological types, and evaluating their distribution after an evaluation task. Most of the new suggested senses satisfy the inclusion criteria we determine. In particular, we advocate for an inclusion of the senses N.DOCUMENT and N.INSTITUTION, which improve NER compatibility.

The extension method can be applied to any wordnet where the synsets are associated to EWN ontological types. Nevertheless, the inclusion criteria might change when dealing with different languages or corpus types. Moreover, the SSI proposed in this article can be applied retroactively to any EWN-aligned synset-annotated corpus.

With regards to adjectives, the backoff A.ALL category still constitutes 40% of the annotated adjectives. In future work, we consider including senses from the GermaNet inventory, and experimenting with data-driven approaches to infer lexical categories for adjectives by means of their relations to other words in wordnets, following the work of Alonge et al. (2000), Mendes (2006), Nimb and Pedersen (2012) and corpus-based approaches like Lapata (2001).

Acknowledgements

We wish to thank Nathan Schneider and Yulia Tsvetkov for their useful comments.

References

Antonietta Alonge, Francesca Bertagna, Nicoletta Calzolari, Adriana Roventini, and Antonio Zampolli. 2000. Encoding information on adjectives in a lexical-semantic net for computational applications. In *Proceedings of the 1st North American chapter of the Association for Computational Linguistics conference*, pages 42–49. Association for Computational Linguistics.

Javier Alvez, Jordi Atserias, Jordi Carrera, Salvador Climent, Egoitz Laparra, Antoni Oliver, and German Rigau. 2008. Complete and consistent annotation of wordnet using the top concept ontology. In *LREC*.

Jørg Asmussen and Jakob Halskov. 2012. The CLARIN DK Reference Corpus. In *Sprogteknologisk Workshop*.

Timothy Baldwin. 2005a. Deep lexical acquisition of verb–particle constructions. *Computer Speech & Language*, 19(4):398–414.

Timothy Baldwin. 2005b. Looking for prepositional verbs in corpus data. In *Proc. of the Second ACL-SIGSEM Workshop on the Linguistic Dimensions of Prepositions and their Use in Computational Linguistics Formalisms and Applications*, pages 180–9.

Gemma Boleda, Sabine Schulte im Walde, and Toni Badia. 2012. Modeling regular polysemy: A study on the semantic classification of Catalan adjectives. *Computational Linguistics*, 38(3):575–616.

Matthias Buch-Kromann, Line Mikkelsen, and Stine Kern Lynge. 2003. Danish dependency treebank. In *TLT*.

Paula Chesley, Bruce Vincent, Li Xu, and Rohini K Srihari. 2006. Using verbs and adjectives to automatically classify blog sentiment. *Training*, 580(263):233.

Nancy Chinchor and Patricia Robinson. 1997. Muc-7 named entity task definition. In *Proceedings of the 7th Conference on Message Understanding*, page 29.

Massimiliano Ciaramita and Yasemin Altun. 2006. Broad-coverage sense disambiguation and information extraction with a supersense sequence tagger. In *Proc. of Proceedings of EMNLP*, pages 594–602, Sydney, Australia, July.

Massimiliano Ciaramita and Mark Johnson. 2003. Supersense tagging of unknown nouns in WordNet. In *Proceedings of the 2003 conference on Empirical methods in natural language processing*, pages 168–175. Association for Computational Linguistics.

Christiane Fellbaum. 1990. English verbs as a semantic net. *International Journal of Lexicography*, 3(4):278–301.

Jane Grimshaw. 1990. *Argument structure*. the MIT Press.

Derek Gross and Katherine J Miller. 1990. Adjectives in wordnet. *International Journal of Lexicography*, 3(4):265–277.

Birgit Hamp and Helmut Feldweg. 1997. Germanet-a lexical-semantic net for German. In *Proceedings of ACL workshop Automatic Information Extraction and Building of Lexical Semantic Resources for NLP Applications*, pages 9–15. Citeseer.

Christina Hoppermann and Erhard Hinrichs. 2014. Modeling prefix and particle verbs in GermaNet. *Global Wordnet Conference*, page 49.

Rubén Izquierdo, Armando Suárez, and German Rigau. 2009. An empirical study on class-based word sense disambiguation. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics*, pages 389–397. Association for Computational Linguistics.

Anders Johannsen, Dirk Hovy, Héctor Martínez, Barbara Plank, and Anders Søgaard. 2014. More or less supervised supersense tagging of Twitter. In *Lexical and Computational Semantics (*SEM 2014)*.

Karin Kipper, Hoa Trang Dang, Martha Palmer, et al. 2000. Class-based construction of a verb lexicon. In *AAAI/IAAI*, pages 691–696.

Karin Kipper, Anna Korhonen, Neville Ryant, and Martha Palmer. 2006. Extending VerbNet with novel verb classes. In *Proceedings of LREC*.

Mirella Lapata and Chris Brew. 2004. Verb class disambiguation using informative priors. *Computational Linguistics*, 30(1):45–73.

Maria Lapata. 2001. A corpus-based account of regular polysemy: The case of context-sensitive adjectives. In *Proceedings of the second meeting of the North American Chapter of the Association for Computational Linguistics on Language technologies*, pages 1–8. Association for Computational Linguistics.

Alessandro Lenci, Nuria Bel, Federica Busa, Nicoletta Calzolari, Elisabetta Gola, Monica Monachini, Antoine Ogonowski, Ivonne Peters, Wim Peters, Nilda Ruimy, et al. 2000. Simple: A general framework for the development of multilingual lexicons. *International Journal of Lexicography*, 13(4):249–263.

Beth Levin. 1993. *English verb classes and alternations: A preliminary investigation*. University of Chicago press.

Oier Lopez de Lacalle and Eneko Agirre. 2015. Crowdsourced word sense annotations and difficult words and examples. *IWCS*.

Héctor Martínez Alonso, Anders Johannsen, Anders Søgaard, Sussi Olsen, Anna Braasch, Sanni Nimb, Nicolai Hartvig Sørensen, and Bolette Sandford Pedersen. 2015. Supersense tagging for danish. In *Nodalida*.

Sara Mendes. 2006. Adjectives in wordnet.pt. In *Proceedings of the GWA 2006–Global WordNet Association Conference*.

George A Miller. 1990. Nouns in wordnet: a lexical inheritance system. *International journal of Lexicography*, 3(4):245–264.

Gregory Leo Murphy and Jane M Andrew. 1993. The conceptual basis of antonymy and synonymy in adjectives. *Journal of memory and language*, 32(3):301–319.

David Nadeau and Satoshi Sekine. 2007. A survey of named entity recognition and classification. *Lingvisticae Investigationes*, 30(1):3–26.

Sanni Nimb and Bolette Sandford Pedersen. 2012. Towards a richer wordnet representation of properties–exploiting semantic and thematic information from thesauri. *LREC 2012*, pages 3452–3456.

Sussi Olsen, Bolette Sandford Pedersen, Héctor Martínez Alonso, and Anders Johannsen. 2015. Coarse-grained sense annotation of Danish across textual domains. In *NODALIDA*.

Rebecca J Passonneau, Ansaf Salleb-Aouissi, and Nancy Ide. 2009. Making sense of word sense variation. In *Proceedings of the Workshop on Semantic Evaluations: Recent Achievements and Future Directions*. Association for Computational Linguistics.

Bolette Sandford Pedersen, Sanni Nimb, Jørg Asmussen, Nicolai Hartvig Sørensen, Lars Trap-Jensen, and Henrik Lorentzen. 2009. Dannet: the challenge of compiling a wordnet for Danish by reusing a monolingual dictionary. *Language resources and evaluation*, 43(3):269–299.

Wim Peters, Ivonne Peters, and Piek Vossen. 1998. Automatic sense clustering in EuroWordNet. In *LREC*. Paris: ELRA.

James Pustejovsky. 1991. The generative lexicon. *Computational linguistics*, 17(4):409–441.

Nathan Schneider and Noah A Smith. 2015. A corpus and model integrating multiword expressions and supersenses. *Proc. of NAACL-HLT. Denver, Colorado, USA. To appear.*

Nathan Schneider, Vivek Srikumar, Jena D. Hwang, and Martha Palmer. 2015. A hierarchy with, of, and for preposition supersenses. In *Proc. of The 9th Linguistic Annotation Workshop*, pages 112–123, Denver, Colorado, USA, June.

Frédérique Segond, Anne Schiller, Gregory Grefenstette, and Jean-Pierre Chanod. 1997. An experiment in semantic tagging using hidden Markov model tagging. In Piek Vossen, Geert Adriaens, Nicoletta Calzolari, Antonio Sanfilippo, and Yorick Wilks, editors, *Automatic Information Extraction and Building of Lexical Semantic Resources for NLP Applications: ACL/EACL-97 Workshop Proceedings*, pages 78–81, Madrid, Spain, July.

Anders Søgaard, Barbara Plank, and Héctor Martínez Alonso. 2015. Using frame semantics for knowledge extraction from Twitter. In *AAAI*.

Fabian M Suchanek, Gjergji Kasneci, and Gerhard Weikum. 2008. Yago: A large ontology from wikipedia and wordnet. *Web Semantics: Science, Services and Agents on the World Wide Web*, 6(3):203–217.

Leonard Talmy. 1985. Lexicalization patterns: Semantic structure in lexical forms. *Language typology and syntactic description*, 3:57–149.

Erik F Tjong Kim Sang and Fien De Meulder. 2003. Introduction to the CoNLL-2003 shared task: Language-independent named entity recognition. In *In CoNLL*.

Yulia Tsvetkov, Leonid Boytsov, Anatole Gershman, Eric Nyberg, and Chris Dyer. 2014a. Metaphor detection with cross-lingual model transfer. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL'14)*.

Yulia Tsvetkov, Nathan Schneider, Dirk Hovy, Archna Bhatia, Manaal Faruqui, and Chris Dyer. 2014b. Augmenting English adjective senses with supersenses. In *Proc. of LREC*.

Piek Vossen, Laura Bloksma, Horacio Rodriguez, Salvador Climent, Nicoletta Calzolari, Adriana Roventini, Francesca Bertagna, Antonietta Alonge, and Wim Peters. 1998. The EuroWordNet base concepts and top ontology. *Deliverable D017 D*, 34:D036.

Fei Wu and Daniel S Weld. 2008. Automatically refining the wikipedia infobox ontology. In *Proceedings of the 17th international conference on World Wide Web*, pages 635–644. ACM.