Lexical Representation of Agentive Nominal Compounds in French and Swedish

Maria Rosenberg

Stockholm University, maria.rosenberg @fraita.su.se

Abstract. This study addresses the lexical representation of French VN and Swedish NV-*are* agentive nominal compounds. The objective is to examine their semantic structure and output meaning. The analysis shows that, as a result of their semantic structure, the compounds group into some major output meanings. Most frequently, the N constituent corresponds to an Undergoer in the argument structure of the V constituent, and the compound displays an Actor role, which more precisely denote entities such as Persons, Animals, Plants, Impersonals, Instruments or Locatives, specified in the Telic role in the Qualia. We propose that the Agentive role can be left unspecified with regard to action modality. In conclusion, our study proposes a unified semantic account of the French and Swedish compounds, which can have applications for NLP systems, particularly for disambiguation and machine translation tasks.

Keywords. Agentive nominal compounds, Actor, Undergoer, semantic structure, lexical representation, Generative Lexicon, telic, disambiguation

1 Introduction

This study addresses the semantics of French and Swedish agentive nominal compounds that contain an N and a V constituent, manifesting an argumental relation. French has only one such compound type, which has exocentric structure [1][1]:

- $[VN_y]_{Nx}$: *porte-drapeau* 'bear-flag=flag bearer'

Table 1 shows the initial data for our study, which aimed to localize Swedish correspondents of French VN compounds. By going through four bilingual French-Swedish dictionaries, we attested 432 French nominal VN compounds. Among these, 229 were rendered by four Swedish compound types. The remaining cases corresponded mainly to simple words or syntactic phrases. Apart from the data in Table 1, our data draws mainly from dictionaries (*TLFi* and *SAOB*) and the Internet. We support our analysis by a restricted sample of representative examples.

[1] Romance VN compounds are also analyzed as left-headed: a nominal zero suffix adds to the V [2], or as right-headed: a nominal zero suffix adds to the compound, considered as a VP, [3].

Special issue: Natural Language Processing and its Applications.
Research in Computing Science 46, 2010, pp. 3-15

Received 24/11/09
Accepted 16/01/10
Final version 08/03/10

Table 1. French VN compounds and their corresponding Swedish compounds in four bilingual dictionaries.

Compound type	*n*
VN (fra)	432
NV-*are* (swe)	108
NV-*a* (swe)	16
NV (swe)	54
VN (swe)	51

According to Table 1, the Swedish NV-*are* compound is the most frequent counterpart. Hence, we focus solely on this Swedish compound in this study. Swedish NV-*are* compounds are right-headed and also called synthetic, since they involve both compounding and derivation. Their formation can be considered as a process of conflation, uniting two templates $[NV]_V$ and $[V\ are]_N$ into a unified productive template $[[NV]\ are]_N$ (cf. [1]):

- $[[NV]\ are]_N$: *fanbärare* 'flag bearer'

French VN and Swedish NV-*are* compounds give rise to polysemy and sense shifting within the agentive domain. They do not only denote humans, but all sorts of animate entities, e.g. animals, plants and insects, as well as artefacts, e.g. instruments, impersonal agents and locatives. They can also refer to places, events and results. Thus, morphologically, VN and NV-*are* correspond to two constructions, which have several underlying semantic structures.

Our main objective is to examine the lexical representation of the compounds. We explore the semantic roles of the N constituents and the semantic characteristics of the V constituents, as well as their semantic structures. Note that both the N constituent and the compound fulfil different roles in the argument structure of the V. The role displayed by the compound corresponds to its output meaning. Despite the formal differences of French and Swedish compounds, we aim at a unified semantic account. Moreover, we discuss the importance of action modality as a component in the lexical representation of agentive compounds. Automatic analyses of nominal compounds constitute an intriguing question within NLP. In order for our study to have some predictive power, we attempt to relate the semantic structures and output meanings to productivity and frequency. At the present, we are setting the ground for a future implementation of regular lexical morphology principles in a machine translation (MT) system.

Section 2 addresses the morphological context. Section 3 deals with the semantic characteristics of the constituents within the compounds. In section 4, we analyze the semantic structures and the output meanings of the compounds. Section 5 discusses the notion of action modality. In Section 6, we propose GL representations for the three most frequent cases. Section 7 discusses potential applications within NLP, and Section 8 contains a conclusion.

2 Compounds in Morphology

The morphological approach is lexeme-based, and adheres to Construction Morphology, being elaborated by Booij (e.g. [1]). A compound is defined as a sequence which cannot be generated otherwise than by morphological rules. Hence, it does not have syntactic structure [4]. This criterion is valid for French VN and Swedish NV-*are* compounds. Interaction between syntax and lexicon is however tolerated: the lexical rules may make use of syntactic information [5]. "Word-formation patterns can be seen as abstract schemas that generalize over sets of existing complex words with a systematic correlation between form and meaning" [1]. The generalizations are expressed in the lexicon by assuming intermediate levels of abstractions between the most general schema and individual existent compounds. Hence, the lexicon is hierarchically structured. The morphology combines three aspects of complex words: phonological form, formal structure and meaning. The architecture of grammar is tripartite and parallel [6]. Abstract schemas coexist with their individual instantiations in the lexicon. Thus, outputs of productive rules can also be listed [1].

3 Semantic Characteristics of the Compounds

3.1 Argument Structure and Semantic Roles of the N Constituent

Agentive compounds contain an argumental relation between the V and N constituents. According to our analysis, the N constituent can correspond, more or less, to all four types of arguments, distinguished in the Generative Lexicon (GL) framework [7]:

- **True arguments**: *ouvre-boîte* 'open-can=can opener', *burköppnare* 'can opener'.
- **Default arguments**: *claque-soif* 'die-thirst=person dying of thirst', *cuit-vapeur* 'boil-steam=steamer', *ångkokare*, 'steam+boiler=steamer' *betonggjutare* 'concrete caster'.
- **Shadow arguments**: *marche-pied* 'march-foot=step, running board', *bensparkare* 'leg kicker'.
- **True adjuncts**: *réveille-matin* 'wake-morning=alarm clock', *trädkrypare* 'tree+crawler=bird'.

In other terms, the semantic roles of the N constituents can correspond to Agent (*croque-madame* 'crunch-madam=toast'), Patient[2] (*ouvre-boîte* 'open-can=can opener'), Theme (*hatthängare* 'hathanger=hat-rack'), Place (*bordslöpare* 'table+runner=cloth'), Time (*dagdrömmare* 'day dreamer'), Instrument/Manner (*cuit-vapeur* 'steam+boiler=steamer', *fotvandrare* 'footrambler', Cause (*claque-soif* 'die-thirst=someone dying of thirst', *sorgedrickare* 'grief+drinker') or Goal (*cherche-pain* 'search-bread=beggar', *målsökare* 'target seeker').

[2] Patient corresponds to an entity, internally affected by the event expressed by the V, whereas Theme corresponds to an entity in motion, in change or being located [8].

However, an overwhelming majority of the French VN, 96 % (415/432), and the Swedish NV-*are* compounds, 97% (105/108), in our data, contain an N which is a direct object of a transitive V. Furthermore, about 73 % (79/108) of the Swedish NV-*are* counterparts of the French VN compounds contain semantically similar lexical units, such as *allume-gaz* 'light-gas=gas lighter' vs. *gaständare* 'gas lighter'. Hence, a MT system could benefit from an implementation of these facts (cf. section 7)

3.2 The Four Classes of *Aktionsart*

The four *Aktionsarten* [9] can occur within the French and Swedish compounds. The state reading is rare, but not unproductive. New formations arise quite easily, such as *godisälskare* 'candy lover'.

- **Activity**: *traîne-nuit* 'loaf-night', *dagdrivare* 'day loafer'.
- **Accomplishment**: *presse-citron* 'squeeze-lemon', *pennvässare* 'pencil sharpener'.
- **Achievement**: *presse-bouton* 'push-button', *cigarrtändare* 'cigar lighter'.
- **State**: *songe-malice* 'think-malice=someone who plots to evil', *vinkännare* 'wine+knower=connoisseur of wine'.

3.3 Unaccusative and Unergative Verbs

We see that both unaccusative and unergative readings of intransitive verbs can be attested within the compounds [10], [11]:

- **Unaccusative**: *caille-lait* 'clot-milk=plant', *oljedroppare* 'oil dripper'.
- **Unergative**: *trotte-bébé* 'baby walker', *hundpromenerare* 'dog+walker=person who takes the dog out for a walk'.

4 Semantic Structures and Output Meanings of the Compounds

The output meaning of French VN and Swedish NV-*are* compounds is taken to be a function of the meanings of their constituents [6]. The agentive compounds themselves display a role in the argument structure of the V (cf. [12] for French VN compounds). The N constituent can be classified for semantic macro-role, Actor or Undergoer [8] (correspond more or less to Proto-Agent and Proto-Patient [13]). In general, the compound corresponds to the Actor (including thematic roles such as Agent, Instrument and Experiencer), and its N constituent to an Undergoer (comprising roles such as Patient, Theme, Source and Recipient) of a transitive V. In order to come up with a more fine-grained semantic analysis, we split up the Actor interpretation into Actors corresponding to first arguments, Instruments, Locatives and Causatives.

We propose that the construction itself links to the output meaning (N_3). The structure of French VN compounds corresponds to $[V_1N_2]_{N3}$ in our analysis. The same proposal is made for Swedish NV-*are* compounds. Instead of linking the Actor

interpretation to the suffix, it links to the entire construction $[N_1V_2\text{-are}]_{N3}$. Thus, the meaning of the compound is the output of its semantic structure. We assume French VN and Swedish NV-*are* compounds to have similar semantic structures, their exocentricity or endocentricity, as well as the order between the V and N constituents, are of minor importance. Two implications follow from our proposal: the formal structure, not the *-are* suffix, is polysemous; null elements are not stipulated. We adopt Jackendoff's framework [6] to account for the semantic structures of the compounds. Table 2 shows the frequency of the output meanings of French VN and Swedish NV-*are* compounds in the initial data. The most frequent cases are Instrument, Agent and Instrumental Locative. They account for more than 90 % of all cases. This figure is confirmed for a collection of 1075 French VN compounds drawn from *TLFi* [14], but further data needs to be added for Swedish NV-*are* compounds.

Table 2. Output meanings of French VN and Swedish NV-*are* compounds in the initial data.

	ACTOR				UND	PLACE	EV	RES	*n*
	Arg1	INSTR	LOC	CAUS					
VN (fra)	128	193	84	3	2	1	18	3	432
	30%	45%	19%	0.7%	0.5%	0.2%	4%	0.7%	
NV-*are* (swe)	40	47	20	1					108
	37%	44%	19%	0.9%					

4.1 Actor is the First Argument

In the Actor interpretation, where the compound corresponds to the first argument of the V, we find compounds such as *porte-drapeau* or *fanbärare*, both 'flag bearer': 'a flag bearer bears a flag' (cf. 1-2). In some cases, the V is intransitive, and the N constituent displays a Place role (cf. 3-4). The compounds denote not only human agents, but also animals (cf. 3), plants, impersonals (cf. 5) (cf. also [15] who relates the Agent polysemy to the Animacy hierarchy). Sometimes, according to the semantics of the V, the compounds manifest an Experiencer role (cf. 6). According to [6], the function PROTECT (X, Y FROM Z) creates two groups of compounds, 'N2 protects N1' (cf. 7-8) and 'N2 protects from N1' (cf. 9-10), which denote some sort of disposal. According to Lieber "verbs which take more than one obligatory internal argument (e.g., *put*) [i.e. ditransitives] cannot form the base of synthetic compounds" [16]. This claim does not seem to be an absolute restriction, in any case not for French and Swedish (cf. also 17-18 in sub-section 4.3).

1. $[\text{porte-}_1\text{drapeau}_2]_3 = \text{PERSON}_3{}^{\alpha}$; $[\text{BEAR}_1\ (\alpha, \text{FLAG}_2)]$
2. $[\text{fan}_1\text{bär}_2\text{are}]_3 = \text{PERSON}_3{}^{\alpha}$; $[\text{BEAR}_2\ (\alpha, \text{FLAG}_1)]$
3. $[\text{trotte-}_1\text{chemin}_2]_3 = \text{ANIMAL}_3{}^{\alpha}$; $[\text{TROT}_1\ (\alpha, \text{ON ROAD}_2)]$
4. $[\text{kåk}_1\text{far}_2\text{are}]_3 = \text{PERSON}_3{}^{\alpha}$; $[\text{GO}_2\ (\alpha, \text{IN SLAMMER}_1)]$
5. $[\text{lave-}_1\text{vaisselle}_2]_3 = \text{MACHINE}_3{}^{\alpha}$; $[\text{WASH}_1\ (\alpha, \text{DISH}_2)]$
6. $[\text{vin}_1\text{känn}_2\text{are}]_3 = \text{PERSON}_3{}^{\alpha}$; $[\text{KNOW}_2\ (\alpha, \text{WINE}_1)]$
7. $[\text{garde-}_1\text{roue}_2]_3 = \text{DISPOSAL}_3{}^{\alpha}$; $[\text{PROTECT}_1\ (\alpha, \text{WHEEL}_2, \text{FROM INDEF})]$
8. $[\text{blus}_1\text{skydd}_2\text{are}]_3 = \text{DISPOSAL}_3{}^{\alpha}$; $[\text{PROTECT}_2\ (\alpha, \text{BLOUSE}_1, \text{FROM INDEF})]$
9. $[\text{garde-}_1\text{boue}_2]_3 = \text{DISPOSAL}_3{}^{\alpha}$; $[\text{PROTECT}_1\ (\alpha, \text{INDEF, FROM MUD}_2)]$

10. [blixt$_1$skydd$_2$are]$_3$ = DISPOSAL$_3^{\alpha}$; [PROTECT$_2$ (α, INDEF, FROM LIGHTNING$_1$)]

4.2 Instrument

Some Instrument denoting compounds are *ouvre-boîte* or *burköppnare*, both 'can opener': 'one opens a can with a can opener', or *casse-noix* or *nötknäppare*, both 'nutcracker'. This meaning is the most productive one in both French and Swedish.

11. [ouvre-$_1$boîte$_2$]$_3$ = INSTR$_3^{\alpha}$; [OPEN$_1$ (INDEF, CAN$_2$, WITH α)]
12. [burk$_1$öppn$_2$are]$_3$ = INSTR$_3^{\alpha}$; [OPEN$_2$ (INDEF, CAN$_1$, WITH α)]
13. [casse-$_1$noix$_2$]$_3$ = INSTR$_3^{\alpha}$; [CRACK$_1$ (INDEF, NUT$_2$, WITH α)]
14. [nöt$_1$knäpp$_2$are]$_3$ = INSTR$_3^{\alpha}$; [CRACK$_2$ (INDEF, NUT$_1$, WITH α)]

4.3 Locative

French VN and Swedish NV-*are* compounds do quite frequently have a Locative interpretation. It is close to the Instrument meaning, but instead of denoting something that one does things with, the compound denotes a location: 'one burns incense in a *brûle-parfum* 'censer'' (cf. 15-16) or 'one hangs saucepans on a saucepan hanger' (cf. 17-18).

15. [brûle-$_1$parfum$_2$]$_3$ = LOC$_3^{\alpha}$; [BURN$_1$ (INDEF, INCENSE$_2$, IN α)]
16. [kaffe$_1$bränn$_2$are]$_3$ = LOC$_3^{\alpha}$; [BURN$_2$ (INDEF, COFFEE$_1$, IN α)]
17. [accroche-$_1$casseroles$_2$]$_3$ = LOC$_3^{\alpha}$; [HANG$_1$ (INDEF, SAUCEPAN$_2$, ON α)]
18. [kastrull$_1$häng$_2$are]$_3$ = LOC$_3^{\alpha}$; [HANG$_2$ (INDEF, SAUCEPAN$_1$, ON α)]

4.4 Causative

Some of the rare French VN compounds that accept unaccusative and unergative Vs receive a reading involving a causative relation. We assume a same semantic structure for both cases: an additional argument (the causer or Actor) adds to the V, and the N is interpreted as an Undergoer (not acting entirely volitionally) of the V (cf. [17]). For example, *trotte-bébé* 'toddle-baby=baby walker', is a device that makes the baby toddle. In Swedish, *folkförödare* 'people+devastater=tuberculosis' involves a causative relation.

19. [trotte-$_1$bébé$_2$]$_3$= DEVICE$_3^{\alpha}$; [CAUSE (α (TODDLE$_1$ (BABY$_2$)))]
20. [folk$_1$föröd$_2$are]$_3$ = DISEASE$_3^{\alpha}$; [CAUSE (α (DEVASTATE$_2$ (PEOPLE$_1$)))]

4.5 Undergoer

Exceptionally, a few French VN compounds have an Undergoer interpretation, in which the N constituent, instead, is an Actor. This case is thus the opposite of the Agent case in sub-section 4.1. For example, *croque-monsieur* 'crunch-sir=toast (that

the sir crunches)', or *pique-poule* (normally spelled as *picpoul*) 'pick-hen=grape (picked by hens)'. This meaning is unproductive in contemporary French, and seems to be ruled out for Swedish NV-*are* compounds.

21. [croque-$_1$monsieur$_2$]$_3$= UND$_3^{\alpha}$; [CRUNCH$_1$ (SIR$_2$, α)]

4.6 Place and Event

Apart from the output meanings above, French VN compounds can denote the place, where the event expressed takes place: *coupe-gorge* 'cut-throat=dangerous place where one risks having one's throat cut'. They are often toponyms, such as *Chante-merle* 'sing-blackbird=a place where the blackbirds sing'. We have not attested any Swedish NV-*are* compound with a Place meaning (cf. *diner* in English).

22. [coupe-$_1$gorge$_2$]$_3$= PLACE$_3$; [LOC (CUT$_1$ (INDEF, THROAT$_2$))]
23. [Chante-$_1$merle$_2$]$_3$= PLACE$_3$; [LOC (SING$_1$ (BLACKBIRD$_2$))]

In addition, French VN and Swedish NV-*are* compounds can denote the event itself expressed by the compound, such as *höftrullare* 'hip roller=rolling the hip'. Some of the compounds with an Event meaning can, according to context, have an additional result interpretation, e.g. *baise-main* kiss-hand 'the act of kissing a hand' vs. 'hand-kiss'.

24. [höft$_1$rull$_2$are]$_3$ = EVENT$_3$; ROLL$_1$ (HIP$_2$)
25. [baise-$_1$main$_2$]$_3$= EVENT$_3$; KISS$_1$ (INDEF, HAND$_2$)

The Place and Event cases do not involve any linking variable. The compound's output meaning does not correspond to a participant in the argument structure of the V; the N constituent can either be an Actor of a V, taking one argument, or an Undergoer of a V, taking two arguments.

5 Action Modality

In [18] a distinction is made between event and non-event English *-er* nominals, corresponding more or less to the distinction between stage-level and individual-level nominals [7], [19]. Inheritance of complement and argument structure correlates with the event interpretation, whereas instruments and occupations, which do not presuppose the existence of an event, typically are non-events. Busa [20], instead, claims that all agentive nominals are best characterized in terms of events, and distinguishes between a changeable property for stage-level nominals, encoded as an Agentive role (cf. 26), and a persistent property for individual-level nominals, encoded as a Telic role (cf. 27):

26. passenger

QUALIA = FORMAL = person
AGENTIVE = travel on vehicle

27. smoker

QUALIA = FORMAL = person
TELIC = smoke

Moreover, Busa [20] argues that state predicates of individual-level nominals can also encode for an agentive role, such as Habit for *smoker* or Ability for *violinist*:

28. violinist

QUALIA = FORMAL = person
TELIC = play violin
AGENTIVE = ability to play violin

29. smoker

QUALIA = FORMAL = person
TELIC = smoke
AGENTIVE = habit to smoke

Jackendoff [6], referring to [21], emphasizes that action modality is an important component for the interpretation and lexical representation of agentive nominals, and not only a matter of pragmatics. There are five major types:

- Current (e.g. *gâte-fête, festförstörare*, 'party trasher')
- Ability (e.g. *gobe-mouches, flugsnappare* 'fly catcher')
- Habit (e.g. *rabat-joie, glädjedödare* 'killjoy')
- Occupation (e.g. *croque-mort, likbärare* 'pall bearer')
- Proper function (*ouvre-boîte, burköppnare* 'can opener')

The Current modality refers to a specific activity on a specific occasion. It concerns stage-level nominals and encodes as an Agentive role. Ability presupposes a potential event (may or may not occur), whilst Habit presupposes repetitive events. Occupation regards persons, practicing the profession indicated by the compound. Proper function concerns objects, and is true irrespectively of actual situations. Thus, the last four modalities involve state predicates and encode as a Telic role, but could additionally encode for an Agentive role [20].

In Table 3, we relate the semantic structures within French VN and Swedish NV-*are* compounds to action modalities. We see that only Proper Function is relevant for objects, all other modalities concern Actors. None of the modalities are relevant for compounds with Place or Event meanings. According to our data, Current (stage-level interpretation) is rarely lexicalized among French VN and Swedish NV-*are* compounds.

Table 3. Semantic structures in relation to action modalities

ACTOR (Arg1)	Current	Habit	Ability	Occupation
CAUSATIVE	Current	Habit	Ability	?
INSTRUMENT	Proper function			
LOCATIVE	Proper function			
PLACE	?			
EVENT	?			

In relation to the general notion of modality, action modality is normally labelled as dynamic, which can be abilitive or volitive [22]. Nuyts [23] proposes dynamic modality to be a subcategory of quantificational aspect, since notions such as "ability/potential" and "need" are semantically similar to notions such as "iterative", "habitual" and "generic". Furthermore, action modality is not overtly linguistically coded and does not affect the lexical content of the verb stem [24]. It can be lexicalized and does not depend solely on context for its interpretation. Hence, action modality, which cannot be defined as an attitudinal expression, seems to be a sort of objective modality [25]. In our opinion, we cannot really see the need for this notion: we propose that the Agentive can be left underspecified for action modality, and that only the Telic is important for the lexical representation of agentive nominal compounds in French and Swedish.

6 GL Representations of the Most Frequent Cases

In order for our study to have some predictive power and importance for NLP systems, we focus on the lexical representation of the three most frequent cases, namely Actor (where the compound is the Arg1 of the V), Instrument and Locative, in which the N constituent is an Undergoer and the compound an Actor in relation to the V. Another possible analysis, different from ours, would be to consider the Actor interpretation as a case of lexical underspecification [26]. The other semantic structures and output meanings, some of them unproductive, are marginal and can probably be exhausted. Nevertheless, instead of proposing a single lexical rule, with a common denominator (cf. [17] for French VN compounds), we propose different lexical representations. The output meanings of the compounds are assumed to be specified in the Type structure, and their internal semantic structure in the Telic role. We do not assume an "instrumental subject"-interpretation of compounds with Instrument or Locative meanings: '*a can opener opens can' or '*a clothes hanger hangs clothes' are not well-formed, in our opinion. Instead, we introduce a default argument, preferably a human agent (not a user *w*, cf. [21]). We use a simplified form of the GL [7], [27], and omit for example Constitutive and Agentive in the Qualia structure.

30. *porte-drapeau, fanbärare* 'flag bearer'

$$\begin{bmatrix} \text{TYPESTR} = & [\,\text{ARG1} = \text{x: human}\,] \\ \text{ARGSTR} = & [\text{D-ARG1} = \text{y: flag}\,] \\ \text{EVENTSTR} = & [\text{D-E1} = \text{e: process}\,] \\ \text{QUALIA} = & \begin{bmatrix} \text{FORMAL} = \text{x} \\ \text{TELIC} = \text{bear\_flag\_act (e, x, y)} \end{bmatrix} \end{bmatrix}$$

31. *ouvre-boîte, burköppnare* 'can opener'

$$\begin{bmatrix} \text{TYPESTR} = & [\,\text{ARG1} = \text{z: artefact\_instrument}\,] \\ \text{ARGSTR} = & \begin{bmatrix} \text{D-ARG1} = \text{x: human} \\ \text{D-ARG2} = \text{y: can} \end{bmatrix} \\ \text{EVENTSTR} = & [\text{D-E1} = \text{e: process}\,] \\ \text{QUALIA} = & \begin{bmatrix} \text{FORMAL} = \text{z} \\ \text{TELIC} = \text{open\_can\_act (e, x, y, with z)} \end{bmatrix} \end{bmatrix}$$

32. *accroche-casseroles, kastrullhängare* 'saucepan hanger'

$$\begin{bmatrix} \text{TYPESTR} = & [\text{ARG1} = \text{z: artefact\_locative}] \\ \text{ARGSTR} = & \begin{bmatrix} \text{D-ARG1} = \text{x: human} \\ \text{D-ARG2} = \text{y: saucepan} \end{bmatrix} \\ \text{EVENTSTR} = & [\text{D-E1} = \text{e: process}] \\ \text{QUALIA} = & \begin{bmatrix} \text{FORMAL} = \text{z} \\ \text{TELIC} = \text{hang\_saucepan\_act (e, x, y, on z)} \end{bmatrix} \end{bmatrix}$$

Note that *x* in the representations (30-32) can be filled with any entity able to display an Actor role. Likewise, *y* can be filled with any entity manifesting an Undergoer role. The events can also be of different types. Our data seems to indicate that intransitive Vs and N constituents with Place or Time meanings (roles displayed by adjuncts in syntax) occur especially in compounds with Actor (Arg1) meanings. In sum, the specification of arguments and predicate structures in the Qualia is important for the analysis of compounds: those included here are all linked to the Telic. Furthermore, phrase structure schemes could be used to account for their compounding (cf. [27]).

7 Discussion

Our analysis of compounds is domain independent, and aims at general semantic structures (cf. [28]), supposed to be lexicalized and more or less productive. Through knowledge about productive semantic patterns, new compounds are created and interpreted [29]. Odd interpretations of compounds are in fact rare [30]. Lapata [31] underlines three problems that compounds still pose for automatic interpretation within NLP: (i) their high productivity implies a need to interpret previously unseen formations; (ii) their internal semantic relation is often implicit; (iii) context and pragmatics have impact on their interpretation.

Contextual information can help to disambiguate unknown compounds of the types included in our study: e.g. subject position in combination with Actor (Arg1) interpretation, "with" in combination with Instruments, and "in" or "on" in combination with Locatives (cf. however [30] for the problematic distinction between agent and instrument at both the morphological and the syntactic level). Since the V constituents in French VN and Swedish NV-*are* compounds cannot always occur as independent Ns in syntax (**porte-*, **häng-/*hängare*), it is not possible to map each of the constituents onto a conceptual representation as is possible for NN root compounds (cf. the systems of [32], [33]). However, a disambiguation algorithm can map the V constituents to their respective verbs and examine distributional properties: e.g. retrieve frequencies of the verb's relation to its objects (verb-argument tuples). In the majority of cases, the N constituent is an internal argument of the (transitive) V constituent. The set of possible interpretations provided by our study enables manual disambiguation of compounds in context, which then can be added to the lexicon.

Our unified account of French VN and Swedish NV-*are* compounds can have relevance for MT or other multi-lingual language processing tasks with regard to Romance and Germanic languages: the GL representation constitutes a neutral platform [27], [32]. Cartoni [34] proposes a prototype of a MT system for handling constructed neologisms, and to which our analysis could be fitted. The first module

checks unknown words with regard to their being potentially constructed or not. If they are, it performs a morphological analysis of their structure and lexeme-bases. The second module generates a possible translation of the analyzed construction. The prototype relies on lexical resources and a set of bilingual Lexeme Formations Rules. The lexeme-bases are checked against the lexical resources and the rules provide information of how to translate them into the target language (e.g. French $V_xN_y \rightarrow$ Swedish $N_yV_x$are: *brise-glace* 'break-ice=icebreaker'$\rightarrow$ *isbrytare*, or alternatively, if the French V constituent corresponds to a lexically established N in Swedish, French $V_xN_y \rightarrow$ Swedish N_yN_x: *appuie-tête* 'rest-head=headrest' $\rightarrow$ *huvudstöd*).

8 Conclusion

This study has attempted to provide a unified account of the complex semantics of French VN and Swedish NV-*are* compounds. We have adopted the frameworks of Jackendoff [6] and GL [7], [27], and been able to find some general semantic structures giving rise to particular output meanings. In the most productive semantic structures, the compounds as well as the N constituents display a role in the argument structure of the V constituent. We assume the Telic role in the Qualia to be most important for their lexical representation. Contrary to the opinion expressed in [20], we suggest that the Agentive role can be left un(der)specified, since it does not add much to their disambiguation or analysis. In conclusion, we hope that our study can have application for NLP systems. Possible applications could be to elaborate a probabilistic algorithm dealing with a disambiguation task for unseen compounds within domain-independent unrestricted text. Our unified account also has relevance for machine translation between French and Swedish, and for multi-lingual language processing with regard to Romance and Germanic languages.

Dictionaries

Fransk-svensk ordbok. (1995). Natur och kultur, Stockholm.
Norstedts fransk-svenska ordbok. (1993). Norstedt, Stockholm.
Norstedts stora fransk-svenska ordbok. (1998). Norstedt, Stockholm.
Norstedts stora svensk-franska ordbok. (1998). Norstedt, Stockholm.
TLFi, Le Trésor de La Langue Française informatisé. http://atilf.atilf.fr/tlf.htm
SAOB, Svenska Akademiens Ordbok. http://g3.spraakdata.gu.se/saob/

References

1. Booij, G.: Compounding and Construction Morphology. In: Lieber, R., Štekauer, P. (eds.) The Oxford Handbook of Compounding, pp. 201-216. Oxford University Press, Oxford (2009)

2. Bisetto, A.: Italian Compounds of the *Accendigas* Type: A Case of Endocentric Formations?. In: Bouillon, P., Estival, D. (eds.) Proceedings of the Workshop on Compound Nouns: Multilingual Aspects of Nominal Composition, pp. 77-87. ISSCO, Geneva (1994)
3. Lieber, R.: Deconstructing Morphology: Word Formation in Syntactic Theory. University of Chicago Press, Chicago (1992)
4. Corbin, D.: Hypothèses sur les frontières de la composition nominale. Cahiers de grammaire 17, 25-55 (1992)
5. Booij, G.: The Grammar of Words. Oxford University Press, Oxford (2005)
6. Jackendoff, R.: Compounding in the Parallel Architecture and Conceptual Semantics. In: Lieber, R., Štekauer, P. (eds.) The Oxford Handbook of Compounding, pp. 105-128. Oxford University Press, Oxford (2009)
7. Pustejovsky, J.: The Generative Lexicon. MIT Press, Cambridge, MA (1995)
8. Van Valin, R. D., Jr.: Semantic Macroroles in Role and Reference Grammar. In: Kailuweit, R., Hummel, M. (eds.) Semantische Rollen, pp. 62-82. Narr, Tübingen (2002)
9. Vendler, Z.: Verbs and Times. The Philosophical Review 66, 143-160 (1957)
10. Burzio, L.: Italian Syntax: A Government-Binding Approach. Reidel, Dordrecht (1986)
11. Perlmutter, D. M.: The Split Morphology Hypothesis. In: Hammond, M., Noonan, M. (eds.) Theoretical Morphology: Approaches in Modern Linguistics, pp. 79-100. Academic Press, San Diego (1988)
12. Fradin, B.: On a Semantically Grounded Difference between Derivations and Compounding. In: Dressler, W. U. *et al.* (eds.) Morphology and its Demarcations, pp. 161-182. Benjamins, Amsterdam/Philadelphia (2005)
13. Dowty, D.: Thematic Proto-Roles and Argument Selection. Language 67, 547-619 (1991)
14. Rosenberg, M.: La formation agentive en français : les composés $[VN/A/Adv/P]_{N/A}$ et les dérivés V-*ant*, V-*eur* et V-*oir*(*e*). PhD Dissertation, Department of French, Italian and Classical Languages, Stockholm University (2008)
15. Dressler, W. U.: Explanation in Natural Morphology, Illustrated with Comparative and Agent-Noun Formation. Linguistics 24, 519-548 (1986)
16. Lieber, R.: Morphology and Lexical Semantics. Cambridge University Press, Cambridge (2004)
17. Roussarie, L.,Villoing, F.: Some Semantic Investigation of the French VN Construction. In: Bouillon, P., Kanzaki, K. (eds.) Proceedings of the Second International Workshop on Generative Approaches to the Lexicon, Geneva, Switzerland (2003).
18. Rappaport Hovav, M., Levin, B.: *-er* Nominals: Implications for the Theory of Argument Structure. In: Stowell, T., Wehrli, E. (eds.) Syntax and the Lexicon. Syntax and Semantics 26, pp. 127-153. Academic Press, San Diego, CA (1992)
19. Carlson, Gregory N. 1977. *Reference to Kinds in English*. PhD Dissertation, University of Massachusetts, Amherst.
20. Busa, F.: The Semantics of Agentive Nominals. In: Saint-Dizier, P. (ed.) Proceedings of ECAI Workshop on Predicative Forms for the Lexicon. Toulouse, France (1996)
21. Busa, F. Compositionality and the Semantics of Nominals. PhD Dissertation, Department of Computer Science, Brandeis University (1996)
22. Palmer, F. R.: Mood and Modality. 2nd ed. Cambridge University Press, Cambridge (2001)
23. Nuyts, J.: The Modal Confusion. In Klinge, A., Müller, H. H. (eds.) Modality: Studies in Form and Function, pp. 5-38. Equinox, London (2005)
24. Bybee, J. L.: Morphology: A Study of the Relation between Meaning and Form. Benjamins, Amsterdam/Philadelphia (1985)
25. Herslund, M.: Subjective and Objective Modality. In: Klinge, A., Müller, H. H. (eds.), Modality: Studies in Form and Function, pp. 39-48. Equinox, London (2005)
26. Pustejovsky, J.: The Semantics of Lexical Underspecification. Folia Linguistica 32, 323-348 (1998)

27. Johnston, M., Busa, F.: Qualia Structure and the Compositional Interpretation of Compounds. In: Viegas, E. (ed.) Breadth and Depth of Semantic Lexicons, pp. 167-189. Kluwer, Dordrecht (1999)
28. Fabre, C.: Interpretation of Nominal Compounds: Combining Domain-Independent and Domain-Specific Information. In: Proceedings of the 16th Conference of Computational Linguistics, pp. 364-369. Copenhagen, Denmark (1996)
29. Ryder, M. E.: Ordered Chaos: The Interpretation of English Noun-Noun Compounds. University of California Press, Berkeley (1994)
30. Isabelle, P.: Another Look at Nominal Compounds. In: Proceedings of the 10th International Conference on Computational Linguistics and the 22nd Annual Meeting of the Association for Computational Linguistics, pp. 509-516. Stanford, California (1984)
31. Lapata, M.: The Disambiguation of Nominalizations. Computational Linguistics 28, 357-38 (2002).
32. McDonald, D. B.: Understanding Noun Compounds. PhD Dissertation, Carnegie Mellon University, Pittsburgh, Pennsylvania (1982)
33. Finin, T.: The Semantic Interpretation of Nominal Compounds. In: Proceedings of First Annual National Conference on Artificial Intelligence, pp. 310-315. Stanford, California (1980)
34. Cartoni, B.: Lexical Morphology in Machine Translation: A Feasibility Study. In: Proceedings of the 12th Conference of the European Chapter of the ACL, pp. 130-38. Association for Computational Linguistics (2009).