Informational Balance
Understanding Romance and Germanic Compounding in a Lexico-Typological Perspective
Høeg Müller, Henrik

Link to publication in CBS Research Portal

Informational balance

Understanding Romance and Germanic compounding in a lexico-typological perspective

Henrik Høeg Müller

Copenhagen Business School (CBS)

Department of International Language Studies and Computational Linguistics (ISV)

1. Introduction

The title of my talk is "Informational balance. Understanding Romance and Germanic Compounding in a lexico-typological perspective". What I basically mean by informational balance is that semantic content is distributed systematically differently between nouns and verbs in the Romance and Germanic languages, and that this distribution is complementary. I shall explain that in detail in a minute, but first I shall introduce you to the problem, which I believe can be explained on the basis of this idea about "informational balance".

2. NP-hypothesis

A frequently observed difference between Germanic and Romance languages is that Germanic languages often use compounding to express what Romance languages convey by a derivational strategy. This means that many simple and derived words in Spanish have compounds as their translational equivalents in Danish, as you can see in 1.

1. Romance simple nouns and derivations vs. Germanic compounds

Simple noun → compound
- berberecho – hjertemusling   [heart-mussel] 'cockle'
- búho – hornugle   [horn-owl] 'horned owl'
- púlpito – prædikestol   [preach-chair] 'pulpit'

Derivation → compound

- escritorio – skrivebord   [write-desk] 'writing desk'
(escribir: 'write')

- dentadura – tandsæt   [tooth-set] 'set of teeth'
(diente: 'tooth')

- petrolero – tankskib   [tank-ship] 'oil tanker'
(petróleo: 'oil')

I believe that this cross-linguistic contrast can be accounted for by means of a lexical-typological hypothesis which builds on the distinction between endo- and exocentric languages, i.e. informational balance or distribution of semantic traits.

2.1. Exocentric vs. endocentric languages

The distinction between exocentric and endocentric languages takes its point of departure in the typological basic assumption that the Romance languages concentrate information in the nominal arguments of the verb, while the verb itself is left relatively underspecified, which prototypically results in a description of the relevant state of affairs as abstract relations between specific entities. Complementary to this distribution of information, the Germanic languages concentrate information in the verb and leave the nominal argument relatively underspecified, which prototypically results in a description of the relevant state of affairs as rather concrete relations between underspecified entities. The theory of endocentric and exocentric languages has been developed by the Danish research group TYPOlex from CBS, and it is comprehensively described in publications by Herslund and Baron, e.g. 2003 and 2005, which are listed in the bibliography.

2.1.1. Verbs

The postulated informational or lexical specificity of the Danish endocentric verbs is based on the insight that these verbs generally lexicalise the semantic component MANNER and, as a result of that, impose selectional restrictions mainly on the subjective argument with which they combine. As illustrated in the left column of (2) below, the Danish expressions corresponding to the Spanish verb *entrar* (enter) are composed by a verb containing the MANNER component and a particle of

directionality *ind* (into), and they adjust semantically to the usual manner of "entering" performed by the subject referents (running, swimming, flying, etc.). On the contrary, the Spanish verb *entrar* (enter) does not contain any information on the manner in which the motion is carried out, and consequently it combines with any subject whose referent can perform a movement, as shown in the right column of (2). This distinction based on MANNER has been known at least since Talmy (1985), but what is essential to understand here is that this MANNER-feature pervades or runs through the whole language.

2. Endocentric and exocentric languages. Verbs.

Endocentric languages [Danish] **Exocentric languages [Spanish]**

• *hunden*	**løb ind**	*el perro*	***entró***	[the dog]
• *fisken*	**svømmede ind**	*el pez*	***entró***	[the fish]
• *fuglen*	**fløj ind**	*el pájaro*	***entró***	[the bird]
• *bilen*	**kørte ind**	*el coche*	***entró***	[the car]
• *skibet*	**sejlede ind**	*el barco*	***entró***	[the ship]
• *manden*	**gik ind**	*el hombre*	***entró***	[the man]

The next examples in (3) show that even in cases where Danish verbs do not contain any clear indication of MANNER – i.e. if they denote some abstract or non-visible activities – there will be a tendency towards introducing the MANNER-component by adding a MANNER-denoting verb in a serial construction, very often as a way of translating the Romance imperfective aspect.

3. Endocentric and exocentric languages. Verbs (continued)

• Descansaba
 Han hvilede sig → Han **lå** og hvilede sig.
 [he rested himself → he lay and rested himself] 'he was resting'

- Leía

  Hun læste → Hun **sad** og læste.

  [she read → she sat and read] 'she was reading'
- Dormían

  De sov → De **lå** og sov.

  [they slept → they lay and slept] 'they were sleeping'
- Comía

  han spiste → han **sad** og spiste.

  [he ate → he sat and ate] 'he was eating'

The examples in (4) illustrate, first, the selectional restrictions imposed by MANNER. Popularly speaking, we had to change verbs in Denmark, when we changes telephones, and turned manual car windows into automatic ones.

4. Endocentric and exocentric languages. Verbs (continued)

Selectional restrictions due to manner
- *Han har drejet/tastet/trykket nummeret → Ha marcado el número*

  [he has rotated/typed/pressed the number] 'he has dialed the number'
- *Rul vinduet ned → baja la ventana*

  'roll the window down'

Free predicative constructions
- *Han græd lommetørklædet vådt → Mojó el pañuelo llorando*

  'he cried the handkerchief wet' → he wetted the handkerchief crying
- *Han strøg tøjet tørt → Secó la ropa planchando*

  'he ironed the clothe dry' → 'he dried the clothe ironing'

Violation of selectional restrictions
- *Edderkoppen sidder/ *står på bordet → La araña está en la mesa*

  'the spider sits/ *stands on the table → the spider is on the table'

The free predicative constructions, intensively studied by Hanne Korzen in Danish and other Germanic languages, are usually not an option in the Romance language, again because of the MANNER verbs, which in these cases are constructed with an alternative argument structure. I mean, we are so eager to express MANNER that we even impose a different argument structure on our MANNER verbs.

The last example I have just added to show you what can happen if we don´t respect, in this case the MANNER specification of the position of a spider. Spiders they don't stand, unless they are frightening huge, they sit in Danish.

In Spanish it is, of course, also possible to specify the manner in which the actions are realised. This can be done via adverbial satellites in the form of prepositional phrases or gerunds or through imperfective, unergative MANNER verbs such as *bailar* (dance) or *nadar* (swim). However, usually the MANNER component is only expressed in Spanish if this aspect of the verbal action is of specific importance or cannot be extrapolated from the situation. In contrast to this, the MANNER component is part of the lexical make-up of most Danish verbs – with some exceptions of course – and therefore Danish verbs are obliged to express the manner, they simply cannot avoid it.

2.1.2. Nouns

As illustrated in (5) and (6) below, the nouns of the two language types show opposite lexicalisation structures to those of the verbs. Exocentric Spanish nouns are generally lexically more contentful and precise than the endocentric ones. Whereas the exocentric Romance nouns tend to lexicalise the semantic component FIGURE, i.e. the shape, dimensionality and structure of the object, endocentric Danish or Germanic nouns nouns are inclined to lexicalise only the component FUNCTION, which, of course, is an inherent abstract feature of any artefact denoting noun. This exclusive focus on the purpose of the object, or non-focus on its form, means that Danish simple nouns in many cases are semantically vague and therefore they may function as denominations on a generic prototype level, i.e. a general hyperonymic level, which for the above mentioned reason of lexicalisation of form in the Romance

languages does not exist in Spanish. When in Danish a level below the general hyporonymic one is needed, this is usually achieved by means of nominal compounds as is shown in (5) and (6). The nouns *vogn* (wagon)*,* an object used for transportation, and *tæppe* (carpet)*,* an object used for decoration or covering things, represent lexicalisation on the family level, while the Romance nouns, because of their lexical specification of outer appearance, must denote subtypes, i.e. objects on a hyponymic level. A predictable consequence of this difference is of course that the Danish nouns *vogn* and *tæppe* cannot be adequately translated into Romance without a proper context.

5. Endocentric and exocentric languages. Nouns

Endocentric languages [Danish]	**Exocentric languages [Spanish]**	
• *tæppe*	[Ø]	lit. transl.
• *senge**tæppe***	*colcha*	[bed –]
• *væg**tæppe***	*tapiz*	[wall –]
• *ægte **tæppe***	*alfombra*	[genuine –]
• *væg-til-væg **tæppe***	*moqueta*	[wall to wall –]
• *teater**tæppe**/ scene**tæppe***	*telón*	[theatre/ stage –]
• *slumre**tæppe***	*manta*	[slumber –]

6. Endocentric and exocentric languages. Nouns (continued)

Endocentric languages [Danish]	**Exocentric languages [Spanish]**	
• *vogn*	[Ø]	lit. transl.
• *personvogn*	*coche*	[person –]
• *lastvogn*	*camión*	[load –]
• *godsvogn*	*vagón*	[goods –]
• *sækkevogn*	*carretilla*	[sack –]
• *hyrevogn*	*taxi*	[hire –]
• *ladvogn*	*plataforma*	[platform –]
• *varevogn*	*furgoneta*	[goods –]

The shell fish and birds of prey taxonomies in (7) illustrate the same phenomenon. Spanish does not have a general hyperonym, and its hyponyms are not necessarily compounds as in Danish, where the hyponym level is predominantly lexicalised on the basis of composite nouns (see Herslund 1997: 31).

7. Taxonomy of shell fish and birds of prey

- 1. generic hyperonym, class   (*skaldyr – marisco*) [shell fish]
- 2. general hyperonym, family   (*musling – Ø*)
- 3. hyponyms, species   (*blå-musling, hjerte-musling, kam-musling, venus-musling – mejillón, berberecho,vieira, almeja*) [mussel, cockle, scallop, clam]

- 1. generic hyperonym, class   (*rovfugl – ave rapaz*) [bird of prey]
- 2. general hyperonym, family   (*falk – Ø*)
- 3. hyponyms, species   (*jagt-falk, lærke-falk, vandre-falk, tårn-falk – gerifalte, alcotán, halcón, cernícalo*) [gerfalcon, hobby, peregrine falcon, kestrel]

Now, maybe you are thinking that to state my case, I have carefully chosen these examples from my personal cabinet of linguistic curiosities or wonders, but I promise you, it is not so. You find these examples in abundance, and the pattern is totally systematic. In (8), to try to convince you, I list a number of examples from different ontological areas, you could say.

8. Examples from different areas

- Professions:
    - **læge →**   *dyr-læge, børne-læge, øjen-læge – veterinario, pediatra, oculista*
    [veterinary, pediatrist, oculist]
- Institutions:
    - **ret →**   *by-ret, lands-ret, højeste-ret – juzgado, audiencia (territorial), tribunal supremo*
    [city court, high court, supreme court]
- Artefacts:
    - **stang →**   *jern-stang, fiske-stang, plejl-stang, – barra, caña (de pescar), biela*
    [bar, rod, connecting rod]
- Plants:
    - **nød →**   *hassel-nød, val-nød, pistacie-nød – avellana, nuez, pistacho*
    [hazelnut, walnut, pistachio nut]
- Animals:
    - **ugle →**   *horn-ugle, slør-ugle, nat-ugle – búho, lechuza, cárabo*
    [horned owl, barn owl, tawny owl]
- Body parts
    - **skæg →**   *fuld-skæg, over-skæg, fip-skæg – barba, bigotes, perilla*
    [(full) beard, moustache, vandyke beard]

2.1.3. Perspectives

The two major word classes, nouns and verbs, are semantically complementary both within the single languages and across language types. Lexical information is organised differently in the Romance and Germanic languages due to the different distribution of semantic content on the two word classes.

The question is in this context, – how do these typological differences affect the issue of compounding?

A consequence of the semantic vagueness of the Danish simple nouns and, accordingly, their lexicalisation on a hyperonymic family level is that in order to designate entities on a hyponymic level Danish must make use of the compounding system. Conceptualising or communicating about concrete species of e.g. shell fish, wagons, carpets, birds of prey, etc. require a word formation system that is capable of specifying the constitution, form, purpose or origin of the object in question and hence create an expression which denotes a subtype. Because of the great demand of such a mechanism, composition is very frequent in Danish and therefore incorporated into the grammatical system as a highly automated morphological word-formation process.

On the contrary, Romance simple nouns are already saturated in a semantic sense, so the Romance languages do not necessarily need, and consequently they have not developed, a full morphological system to deal with this information packaging, or sub-type denotation, task. Either the semantic components are already encapsulated in the simple noun or they use an alternative strategy, namely derivation, to lead the original lexical unit in another semantic direction. So although phrasal composition of the [N prep. N]-type also in the Romance languages is very often a prerequisite for creating subtype denoting lexical expression, it can be regarded as additional to the derivational system and the semantically contentful simple nouns. Therefore composition in the Romance languages is not routinised as part of a morphological system, but has the status of a syntactic devise of a phrasal and complementary nature.

Another reflection that supports the view of a fundamental typological difference between the two language types, and which influences their word-formation systems, is the fact that when the semantically underspecified Danish nouns – such as for example the lexeme *vogn* in (5), which denotes any mobile device for transportation with wheels – occur as heads in compounds, they resemble the semantically

underspecified derivational suffixes of the Romance languages. This point is illustrated with the examples in (9). The analogy between the categories consists in the fact that both the Danish nouns and the Spanish suffixes are semantically vague in the sense that in isolation they do not convey any specific meaning. Both types of elements need to be lexically saturated from outside to obtain full "denotational status". They share the feature of semantic inspecificity, but syntactically the Danish nouns function as heads and the Spanish suffixes as modifiers in relation to the nominal elements with which they unite. The analogy is further corroborated by the fact that the Danish nouns, when functioning as heads, are even reduced prosodically and pronounced with secondary stress.

9. Analogy between Romance derivation and Germanic compounding

Spanish   **Danish**
- *puente – pontaje*   *bro – bropenge*
                                                   [bridge – bridge toll]

- *olmo – olmeda*   *elm – elmelund*
                                                   [elm – elm grove]

- *diente – dentadura*   *tand – tandsæt*
                                                   [tooth – set of teeth]

- *petróleo – petrolero*   *olie – oliehandler*
                                                   [oil – oil dealer]

- *escribir – escritorio*   *skrive – skrivebord*
                                                   [write – writing desk]

- *campana – campanario*   *klokke – klokketårn*
                                                   [bell – belfry]

- *plátano – platanal*   *banan – bananplantage*
                                                   [banana – banana grove]

I think that these data point towards or maybe even substantiates the overall assumption that Danish nouns are semantically imprecise or ambiguous, and that Danish, as a result of that, has developed a nominal compositional system which is a

morphological process invisible to syntax. The Romance languages have more difficulties in combining semantic heavy weight nouns, and for that reason they make comparatively more use of the derivational system or, of course, the phrasal compounding system which is predominantly a syntactic process.

Of course, a number of questions can be raised with respect to this hypothesis, but here I shall only mention a few of the most obvious ones:

(i)   The general characterisation of Danish nouns as being semantically vague is based on a limited number of nouns – this is not a statistical study –, and these nouns are typically equivalent to Spanish simple nouns or derivations. In this connection we must not forget that in many cases there is actually a structural parallelism between Spanish and Danish, cf. e.g. 'barco de guerra/ krigsskib' [war ship], 'granada de mano/ håndgranat' [hand grenade] and 'máquina de escribir/ skrivemaskine' [type writer], where both languages use compositional structures.

(ii) Also, it is not immediately evident that the generalisation made about Danish compounding (or Germanic in general, for that matter) is entirely valid, as heads of Danish compounds cannot immediately be judged semantically vague in all cases. This is especially evident when the head noun instead of an artifact denotes an abstract entity as in 'indkomstskat/Einkommensteuer' (income tax) or 'regeringskonference/Regierungskonferenz' (intergovernmental conference). So the question is how far this generalisation can be justified and in what sort of situations it will prove incorrect.

(iii) In many cases the status of Romance syntagmatic structures is unclear as to whether they should be regarded as compound-like constructions or free syntactic phrase formations. In the Romance languages the compoundedness of phrasal structures is commonly measured by degree of lexicalization, i.e. a semantic concept, and not formal evidence as in the non-English Germanic languages. As the criteria for compounding vary from language type to language type, it is questionable on what basis the comparison is being made.

On this occasion the hypothesis has be introduced and substantiated with a limited number of examples, but there is no doubt that it is an empirical matter requiring more detailed studies and above all statistically significant data.

List of references

- Bally, C. 1932. *Linguistique générale et linguistique française*. Berne: Francke.
- Croft, W. & Cruse, D.A. 2004. *Cognitive Linguistics*. Cambridge: Cambridge University Press.
- Herslund, M. (ed.). 1997. *Det franske sprog. Kapitel I. Grundlag*. Preliminary version. Copenhagen: Copenhagen Business School.
- Herslund, M. & Baron, I. 2003. Language as World View. A Typological Approach to Danish and French. Language and Culture. *Copenhagen Studies in Language* 29. 29-42.
- Herslund, M. & Baron, I. 2005. Langues endocentriques et langues exocentriques. Approche typologique du danois, du français et de l'anglais. In M. Herslund & I. Baron (eds), *Le Génie de la Langue Française. Perspectives typologiques et contrastives*. (Langue française 145) Paris: Larousse. 35–53.
- Korzen, H. 2004. Les attributs indirects en français et en danois: Différences typologiques et problèmes de traduction". In G. Hansen, K. Malmkjaer and D. Gile (eds.), *Claims, Changes and Challenges in Translation Studies*. 73–81.
- Korzen, I. 2008. Determination in endocentric and exocentric languages. With evidence primarily from Danish and Italian. In H. H. Müller & A. Klinge (eds), *Essays on Nominal Determination. From Morphology to Discourse Management*. (Studies in Language Companion Series (SLCS) 99). Amsterdam/Philadelphia: John Benjamins. 69–100.
- Korzen, I. & Lundquist, L. (red.). 2004. *Sprogtypologi og oversættelse. Endocentriske og exocentriske sprog*. Copenhagen: Samfundslitteratur.
- Müller, H.H. 2005. Meaning Construction within Spanish Nominal Syntagmatic Compounds (NSCs). In I. Korzen & P. D'Achille (eds.), *Tipologia Linguistica e Società. ConsiderazioniIinter- e Intralinguistiche..* Firenze: Franco Cesati Editore. 55-76.

- Müller, H.H. 2007. Predicatives inside NPs. In I. Korzen (ed.), *L'Europe : diversité linguistique — unité de pensée? Scolia 22*. I. Strasbourg. S. 63-83.
- Rainer, F. & Varela, S. 1992. Compounding in Spanish. *Rivista di Linguistica* 4(1): 117–142.
- Rosch, E. 1975. Cognitive representation of semantic categories. *Journal of Experimental Psychology*, 104: 192-233.
- Rosch, E. 1978. Principles of categorization. In E. Rosch & B.B. Lloyd (eds.), *Cognition and categorization*. Hillsdale, NJ: Erlbaum. 27–47.
- Talmy, L. 1985. Lexicalization patterns: semantic structure in lexical form. In T. Shopen (ed.), *Language typology and the lexicon,* vol. III: *Grammatical categories and the lexicon.*. Cambridge: Cambridge University Press. 57-149.
- Wierzbicka, A. 1996. *Semantics: primes and universals*. Oxford: Oxford University Press.