

AMBIGUOUS WORDS

MISMATCHES BETWEEN SYNTAX AND THE LEXICON

1. STATE OF THE ART

Consider the contrast in grammaticality between (1) and (2).

- (1) Mary kisses John.
- (2) *Mary laughs John.

The standard account for this contrast is well-known (Gruber 1965, Fillmore 1968, Chomsky 1981, Reinhart 2002 amongst many others). A transitive verb, such as *kiss* in (1), is used in a sentence with both a subject and direct object, whereas an intransitive verb, as *laugh* in (2) only licenses a subject.

More generally, the relation between words and syntactic structures is traditionally considered to be straightforward and one-to-one: transitive verbs occur in transitive sentences, nouns occur in noun phrases, prepositions in prepositional phrases and so forth. Put differently, words have inherent properties, such as the property of being a(n) (in)transitive verb, and these properties match the syntactic context in which these words appear.

The precise implementation of this idea and the additional assumptions vary considerably (see Chomsky 1965, Kiparsky 1973, Halle and Marantz 1993, Pollard and Sag 1994, Doetjes 1997, Grimshaw 2009, Caha 2009, Starke 2011a,...). However, all approaches share the view that there has to be a match between the properties of morphemes and the syntactic structures in which they occur. I will refer to this position as the ‘matching principle’.

(3) Matching Principle

There has to be a match between the inherent properties of morphemes and the syntactic structures in which they occur.

It is this well-established view that is called into question in this project. I focus on data which systematically falsify the matching principle. The main research question is formulated in (4).

(4) Main research question

How should we account for systematic mismatches between the inherent properties of morphemes and the syntactic structures in which they occur?

2. OBJECTIVES

The idea that words match the requirements of the syntactic contexts in which they occur is so straightforward and widespread that it seems almost trivial. However, counter-examples to this principle are very common. In this project I will study three

pervasive sets of data which are highly problematic for the matching principle. These sets include examples of self-reference, ambiguous closed-class morphemes and ambiguous open-class morphemes. They all instantiate cases in which words of a certain category occur in a syntactic context of a different category. I will briefly introduce these data below. (I postpone a thorough discussion of their theoretical relevance till section 3.)

Self-reference is the phenomenon of using a morpheme to refer to itself, as in the example below.

- (5) The *that* on the blackboard is a complementizer.

(5) is an example of self-reference: a word, in this case *that*, is used to refer to itself. This is not a marginal phenomenon. It is productive and common cross-linguistically. As far as I know it occurs universally. (6) is an example from Lebanese Arabic (p.c. Sarah Ouwayda).

- (6) L-**ennou** yalli b-ha-l-masal ma-na daruuriyyeh.
the-**that** that in-this-the-example not-it necessary
‘The **that** that is in this example is not necessary’

Self-reference instantiates a mismatch between the properties of the morpheme and its syntactic context. Consider (5) again. *That* may either be a complementizer or a demonstrative, as shown in (7), but in (5) it is used as a noun, as will be argued below (see section 3).

- (7) a. I said **that** I bought a dog.
b. That dog barked.

More generally, self-reference is a domain in which words of any category can be used as a noun. It would not be very insightful to state that for any morpheme in any language there is a nominal copy in the lexicon. It is therefore inherently problematic for the matching principle and it calls for a theory which accounts for these data.

Ambiguous closed-class morphemes are morphemes which seem to have more than category, where at least one of these categories is a closed-class category. A well-known example is the verb *zijn* ‘to be’, as illustrated below. It can function as a copula, as in (8), and as a tense auxiliary, as in (9).

- (8) Ze **is** slim.
she is smart
‘She is smart.’

- (9) Ze **is** gevallen.
she is fallen
‘She fell.’

One could of course assume that this is a case of accidental homonymy: there are two verbs *zijn* ‘to be’ in Dutch, a copula and a tense auxiliary, and they happen to sound the same. However, this approach cannot account for the systematic nature of this ambiguity. This appears from the fact that the ambiguity of *to be* is pervasive cross-linguistically. (10) and (11) are parallel examples from French.

(10) Elle **est** intelligente.
she is smart
‘She is smart.’

(11) Elle **est** tombée.
she is fallen
‘She fell.’

Rather than assuming homonymy, I take it that the verb *to be* is an ambiguous closed-class morpheme: a single underspecified morpheme fits both syntactic structures (see Starke 2011b). Ambiguous closed-class morphemes are problematic for the matching principle. The matching principle states that there should be a one-to-one relation between the morpheme and the syntactic structure: copulas occur in copulative sentences and tense auxiliaries occur in sentences with a perfect tense. However, given the ambiguity presented above, one wonders whether *to be* is essentially a copula, a tense auxiliary, both or neither. In sum, if closed-class morphemes are ambiguous, they disrupt the one-to-one relation between morphemes and structures which is implied in the matching principle.

Ambiguous open-class morphemes are often referred to as cases of conversion. They involve one and the same open-class morpheme which can adopt more than one category. In (12), for example, the morpheme *Britney* is a proper name, in (13) it is an adjective (see De Clercq 2009).

(12) Britney is pregnant.
(13) That’s so Britney!

Again, one could assume that this is another case of accidental homonymy: a verb *Britney* and an adjective *Britney* both happen to be part of the English lexicon. However, such an approach fails to capture the fact that the phenomenon is productive, in that any proper name can be inserted in the sentence in (13) and function as an adjective. This is shown in (14).

(14) a. That’s so Bieber!
b. That’s so Timberlake!

These examples are once again instances of systematic violations of the matching principle.

Above I have introduced three empirical domains which systematically call into question the validity of the matching principle. They have often been put aside as exceptional and marginal, although they are productive and common. The aim of this

project is to address these mismatches and to incorporate them in the theory. The main research question is repeated in (15).

(15) **Main research question**

How should we account for mismatches between the inherent properties of morphemes and the syntactic structures in which they occur?

In answering this question, I will make the radical move of turning the matching principle on its head. Rather than assume that the properties of the morpheme match the syntactic position, I will assume the opposite, namely that the syntactic structure determines the properties of the morpheme (see Borer 2005ab, 2009abc). If syntax builds a transitive sentence the morpheme in the verbal position will be interpreted as a transitive verb, if syntax builds an adjectival phrase, the morpheme which heads it will be interpreted as an adjective, etc. Put differently, if the morpheme *Britney* is interpreted as an adjective, that happens not because *Britney* is an adjective inherently, but because the syntactic phrase is adjectival inherently. The main research hypothesis is formulated in (16).

(16) **Main research hypothesis**

The syntactic structure determines the properties of the morpheme.

3. METHODOLOGY

In this section I discuss in detail the three empirical domains I presented above. I show that they falsify the widespread matching principle. Furthermore, they are far from marginal. It is therefore not justified to put them aside. They can lead to new insights and an empirically more adequate theory on the syntax-lexicon interface.

3.1 Self-reference

As stated earlier, words and parts of words can be used to refer to themselves, as in (17) (see Harris 1979, Postal 2004, De Belder & van Craenenbroeck 2011 and De Belder 2011).

(17) *That* is a complementizer.

The phenomenon is exceptionally productive. Examples can be found with closed-class morphemes, as in (17), with open-class morphemes, as in (18), with bound morphemes, as in (19), with morphemes from another language, as in (20) and with nonsense words, as in (21).

(18) *Cat* is a noun.

(19) *-ize* is a verbalizing affix.

(20) *Gaard* is etymologically related to *jardin*.

(21) *Shomshom* is not an existing word.

In sum, self-reference can involve simply any linguistic expression. Furthermore, it strips morphemes which are listed with syntactic properties of their properties. The word *that* in (17) is no longer a complementizer; it is a noun. In fact, all self-referring words function as nouns. The nominal nature is indicated by the fact that they can combine with determiners, cardinals and plural marking. Examples are given below.

- (22) The three *thats* on the blackboard are all complementizers.
(23) The three *-izes* on the blackboard have slightly different meanings.

Postal (2004) points out that self-reference has been ignored in the literature. It is not hard to imagine why: the data are an embarrassment for the matching principle. Under the research hypothesis in (16) to be adopted in this project, a theory on self-reference becomes possible, though. If one can show that the syntactic structure of self-reference is regular, one can postulate that this structure assigns the self-referring word its nominal properties. Now observe that the examples indeed adhere to a syntactic pattern: they all behave like proper names. I will illustrate this point by means of two tests. Firstly, proper names do not require a determiner (see (24)), except when they are followed by a restrictive modifier, as in (25).

- (24) New York is portrayed in many novels.
(25) The New York of the 1980s is portrayed in many novels.

The examples in (17) and (22) show that examples of self-reference adhere to exactly the same pattern. Secondly, they cannot function as predicates. First observe that complements of copulas are generally predicates, as in (26).

- (26) Mary is intelligent/a genius/the best dancer in town.

However, when proper names are the complement of a copula, the sentence is equative, as in (27).

- (27) John is Mr. Big.

Predicate sentences and equative sentences differ in their ability to be the complement of *to consider* (Geist 2007). This is shown in (28) and (29).

- (28) Pete considers Mary intelligent/a genius/the best dancer in town.
(29) * Pete considers John Mr. Big.

If self-referring words are proper names we expect that the copulative sentence in (30) cannot be the complement of *to consider*. This is borne out (see (31)).

- (30) *Gaard is jardin.*
(31) * Mary considers *gaard jardin.*

In sum, self-referring examples behave like proper names. However, it would be unreasonable to assume that any combination of sounds in any language is listed as a proper name. It follows that the morphemes which behave like proper name are not marked as such lexically. I therefore hypothesize they just become proper names because the syntactic structure in which they occur forces this particular interpretation. As such, examples of self-reference may support the main hypothesis of this project, which is repeated in (32).

(32) Main research hypothesis

The syntactic structure determines the properties of the morpheme.

The challenge of this section is not to determine the syntax of proper names. It is well-known that noun phrases which behave as proper names have specific structural properties (see Longobardi 1994, Borer 2005a). The challenging side of self-reference is rather its pervasive productivity. The relevant subquestion is formulated in (33).

- (33) How can we account for the productivity of self-reference?

One needs a theory which allows any random combination of sounds to enter the syntactic slot of a proper name. Put differently, the self-referring word should be stripped of all its properties (e.g. it should no longer function as a complementizer) and it should adopt all of its properties from its syntactic environment. In this project I aim to develop such a theory.

3.2 Ambiguous closed-class morphemes

Some closed-class morphemes occur in several syntactic contexts. For example, the Dutch morpheme *wat* is ambiguous. It can function both as a question word, as in (34), and as an indefinite determiner, as in (35) (see Postma 1994, Bennis 1995 and Den Dikken 2002).

- (34) Je hebt wat gegeten.
you have what eaten
'You ate something.'

- (35) Wat heb je gegeten?
what have you eaten
'What did you eat?'

Obviously, these examples are problematic for the matching principle because there is no one-to-one relation between the morpheme and a specific syntactic structure: one morpheme can occur in several structures.

As before, we find that the observed ambiguity is systematic, in that it is not restricted to Dutch. It occurs in Serbo-Croatian as well, as shown in the examples below (p.c. Boban Arsenijevic).

(36) Ako ti šta treba,...
 if you.DAT what need
 'If you need anything...'

(37) Šta si video?
 what Aux seen
 'What did you see?'

The Mandarin morpheme *shenme* shows the same ambiguity (p.c. Xiaoli Dong). Because these examples occur cross-linguistically, an insightful approach might unify these data. Postma (1994) has already argued that the ambiguous morpheme adopts properties from the structure. Depending on the syntactic structure in which the ambiguous morpheme occurs, it is assigned a specific interpretation. Ambiguous closed-class morphemes thus support the main hypothesis, which is repeated in (38).

(38) **Main research hypothesis**

The syntactic structure determines the properties of the morpheme.

The challenge for this subpart of the project is the restrictiveness of the ambiguity. The pronoun *wat* is ambiguous enough to function both as an indefinite determiner and as a question word, but it cannot function as a personal pronoun in Dutch. Furthermore, the Dutch question word *wie* 'who' cannot function as an indefinite (meaning 'someone'). The subquestions of this part of the project are formulated in (39).

- (39) How can we account for the limited ambiguity of some closed-class morphemes?
- What is the range of variation? Can one predict which types of ambiguity are attested?
 - What are the properties of closed-class morphemes such that their ambiguity is restricted?

Other good candidates to be approached as ambiguous closed-class morphemes include quantifiers (see De Belder 2011 and den Dikken 2002 on *heel*, Heim 1982 and Postma 1995), degree words (see Doetjes 1997), various sets of pronouns on top of the one discussed above (Kratzer 2009), the diminutive morpheme (Cinque 2007, De Belder 2011, De Belder et al. 2009), negative polarity items (see Postma 2001 and Den Dikken & Giannakidou 2002) and semi-lexical items (cf. Van Riemsdijk 1998, Corver & Van Riemsdijk 2001, De Belder 2011).

3.3 Ambiguous open-class morphemes

In section 2 I have discussed the fact that proper names can be used as adjectives. In section 3.1 I have shown that any random combination of sounds can function as a proper name. One may get the impression that for open-classes anything goes. If anything can function as a noun, then maybe anything can function as an adjective or

a verb as well. This is clearly not the case. It is, for example, uncommon for adjectives to function as verbs in Dutch. This is shown in (40).

(40) A		V
boos	'angry'	*bozen
leuk	'funny'	*leuken
bang	'afraid'	*bangen
zat	'drunk'	*zatten
mooi	'pretty'	*mooien
stout	'naughty'	*stouten
verstandig	'intelligent'	*verstandigen
armoedig	'shabby'	*armoedigen
verliefd	'in.love'	*verliefden

A theory which is liberal enough to include the observed mismatches between the lexicon and syntax, should be at the same time restricted enough to exclude the unwanted cases as well. This yields the last subquestion of the project:

- (41) How can we account for the limited ambiguity of some open-class morphemes?
- What is the range of variation? Can one predict which types of ambiguity are attested?
 - What are the properties of open-class morphemes such that their ambiguity is restricted?

REFERENCES

Bennis, Hans (1995) The meaning of structure: the *wat voor* construction revisited. In: den Dikken, Marcel & Kees Hengeveld (eds.) *Linguistics in the Netherlands 1995*. Amsterdam: John Benjamins Publishing, 25-37.

Borer, Hagit (2005a) *In Name Only: Structuring Sense vol 1*. Oxford: Oxford University Press.

Borer, Hagit (2005b) *The Normal Course of Events: Structuring Sense vol 2*. Oxford: Oxford University Press.

Borer, Hagit. (2009a). *Roots and categories*. Talk presented at the 19th Colloquium on Generative Grammar, University of the Basque Country, April, 1-3 2009.

Borer, Hagit (2009b) *Categorizing Roots*. Talk presented at Roots, Universität Stuttgart, June 10-12, 2009.

Borer, Hagit (2009c) *Root Bound*. Talk presented at the University of Tromsø, October 19, 2009.

Caha, Pavel (2009) *The Nanosyntax of Case*. Doctoral Dissertation, University of Tromsø.

Chomsky, Noam (1965) *Aspects of the theory of syntax*. Cambridge, MA: The MIT Press.

- Chomsky, Noam (1981) *Lectures on Government and Binding: the Pisa Lectures*. Dordrecht: Foris Publications.
- Cinque, Guglielmo (2007) La natura grammaticale del diminutivo e del vezzeggiativo. In: Maschi Roberta et al. (eds.) *Miscellanea di Studi Linguistici offerti a Laura Vanelli da amici e allievi padovani*. Udine: Forum Editrice, 229-236.
- Corver, Norbert and Henk van Riemsdijk (2001) *Semi-lexical categories. The function of content words and the content of function words*. Berlin: Mouton de Gruyter.
- De Belder, Marijke (2011) *Roots and Affixes. Eliminating lexical categories from syntax*. Utrecht: LOT Publications.
- De Belder, Marijke, Noam Faust and Nicola Lampitelli (2009) *On a derivational and an inflectional diminutive*. Talk presented at North East Linguistic Society 40 (NELS 40), MIT, November 13-15, 2009.
- De Belder, Marijke & Jeroen van Craenenbroeck (2011) *How to merge a root*. Ms., HUBrussel & Utrecht University.
- De Clercq, Karen (2009) *The syntax of denominal proper adjectives*. Talk presented at TIN-dag, February 7, 2009.
- den Dikken, Marcel (2002) *Direct and parasitic polarity item licensing*. *Journal of Comparative Germanic linguistics* 5: 35-66.
- den Dikken, Marcel & Anastasia Giannakidou (2002) *From hell to polarity: Aggressively non-D-linked wh-phrases as polarity items*. *Linguistic Inquiry* 33: 31-61.
- Doetjes, Jenny (1997) *Quantifiers and selection: On the distribution of quantifying expressions in French, Dutch and English*. Doctoral Dissertation, Leiden University.
- Fillmore, Charles J. (1968) The case for case. In: Bach, Emmon and Robert Thomas Harms (eds.) *Universals in Linguistic Theory*. New York: Holt, Rinehart and Winston, 1-88.
- Geist, Ljudmila (2007) *Predication and equation in copular sentences: Russian vs. English*. In: Comorovski, Ileana & Klaus von Heusinger (eds.) *Existence: Semantics and Syntax*. Dordrecht: Springer, 79-105.
- Grimshaw, Jane (2009) *The extended projection*. Ms., Rutgers University.
- Gruber, Jeffrey (1965) *Studies in lexical relations*. Doctoral Dissertation, MIT.
- Halle, Morris & Alec Marantz (1993) *Distributed Morphology and the pieces of inflection*. In: Hale, Kenneth & Samuel Jay Keyser (eds.) *The view from building 20*. Cambridge: MIT Press, 111-176.
- Harris, Zellig (1979) *Mathematical structures of language*. New York: Robert Krieger.
- Heim, Irene (1982) *The semantics of definite and indefinite noun phrases*. Doctoral Dissertation, University of Massachusetts.
- Kiparsky, Paul (1973) *Elsewhere in phonology*. In: Kiparsky, Paul and Stephen R. Anderson (eds.) *A Festschrift for Morris Halle*. New York: Holt, Rinehart and Winston.
- Kratzer, Angelika (2009) *Making a pronoun. Fake indexicals as windows into the properties of pronouns*. *Linguistic Inquiry* 40(2), 187-239.
- Longobardi, Giuseppe (1994) *Reference and proper names: a theory of N-movement in syntax and logical form*. *Linguistic Inquiry* 25:4, 609-665.
- Pollard, Carl and Ivan A. Sag (1994) *Head-driven phrase structure grammar*. Chicago: University of Chicago Press.
- Postal, Paul M. (2004) *Skeptical Linguistic Essays*. Oxford: OUP.
- Postma, Gertjan (1994) *The indefinite reading of WH*. In: Bok-Bennema Reineke & Crit Cremers (eds.) *Linguistics in the Netherlands 1994*. Amsterdam: John Benjamins Publishing, 187-198.
- Postma, Gertjan (1995) *Over de interpretatieve samenhang van de kwantoren heel, veel en al*. In: Zwart, Jan-Wouter en Eric Hoekstra (eds.) *Tabu squibnummer deel 2, Jaargang 25 nummer 4*, 185-190.
- Postma, Gertjan (2001) *Negative polarity and the syntax of taboo*. In: Hoeksema, Jacob et al. (eds.) *Perspectives on negation and polarity items*. Amsterdam & Philadelphia: John Benjamins Publishing, 283-330.
- Reinhart, Tanya (2002) *The theta system: an overview*. *Theoretical linguistics* 28: 229-290.
- Starke, Michal (2011a) *Towards elegant parameters: language variation reduces to the size of lexically stored trees*. Ms., Tromsø University.
- Starke, Michal (2011b) *Nanosyntax*. Lecture notes, Ghent, November 7, 2011 - November 10, 2011.
- Van Riemsdijk, Henk (1998) *Categorial feature magnetism: the endocentricity and distribution of projections*. *Journal of Comparative Germanic Linguistics* 2: 1-48. Williams 1981