A two-rule analysis of measure noun phrases

Dan Flickinger
Stanford University

Francis Bond
NTT Communication Science Labs

Proceedings of the 10th International Conference on
Head-Driven Phrase Structure Grammar

Michigan State University

Stefan Müller (Editor)

2003

Stanford, CA: CSLI Publications

pages 111–121

Flickinger, Dan & Francis Bond. 2003. A two-rule analysis of measure noun phrases. In Stefan Müller (ed.), *Proceedings of the 10th International Conference on Head-Driven Phrase Structure Grammar, Michigan State University*, 111–121. Stanford, CA: CSLI Publications. DOI: 10.21248/hpsg.2003.7. CC-BY 4.0

Abstract

In this paper we present an analysis of English measure noun phrases. Measure noun phrases exhibit both distributional idiosyncrasy, in that they appear in positions normally filled by degree adverbs: *a ten inch long string*; and agreement discord: *ten inches is enough*, *it is ten inch/*inches long*. The analysis introduces one idiosyncratic construction, the Measure Phrase Rule, which links together syntax and inflectional morphology. Combined with existing rules, in particular the Noun-noun Compound Rule, the new rule accounts for the both the distributional and agreement idiosyncrasies. The rule has been implemented and tested in the ERG, a broad-coverage grammar of English. Our analysis supports the position that broad-coverage grammars will necessarily contain both highly schematic and highly idiosyncratic rules.

1 Introduction

One of the central themes within the HPSG framework involves the use of a small number of rule schemata to express generalizations about a wide range of syntactic and semantic phenomena. Pollard and Sag (1994) present a handful of rules and show how they can be employed in combination with a rich lexicon to provide explanatory analyses of much of English syntax. Even seemingly idiosyncratic phenomena like English number expressions (Smith, 1999) and tag questions (Bender and Flickinger, 1999) have been analyzed without requiring the addition of special-purpose rule schemata to the grammar. There are, however, phenomena which do seem to require construction-specific syntactic rules (Sag, 1997), either because the construction itself contributes semantic content, as with noun-noun compounds, or because a phrase is exocentric, exhibiting a syntactic distribution not predictable from its head daughter, as with gerunds (Malouf, 2000) (e.g. *Not hiring Sandy will make Kim unhappy.*). In this paper we present an analysis of another phenomenon in English which requires this kind of special-purpose syntactic machinery, namely measure noun phrases like the ones illustrated in (1):

(1) a. *I need a cord that is ten feet long.*

b. *I need a ten foot long cord.*

c. *This bookcase is two shelves higher than my old one.*

d. *That three gallons was enough to get me home.*

e. *Kim gave me twelve dollars, but I have already spent it/#them.*

The phrases *ten feet* and *two shelves* have the distribution here of degree specifiers like *very* or *much*, yet consist of a numerical expression and a noun, neither of which lexically possess the properties needed to license their appearance as a degree specifier. Further, the noun inflects differently in predicative adjective phrases (1a) and attributive adjective phrases (1b).

Another well known property of measure noun phrases is that, even though their head is plural, they can be modified by singular determiners, and agree with

singular verbs, as in (1d) (e.g., Payne and Huddleston, 2002, 354). In addition, a measure noun phrase can be referred to by a singular pronoun, as in (1e). Bond (2001) points out that such noun phrases are ambiguous. If the noun phrase refers to 12 actual dollar coins or notes, then the noun phrase is a prototypical noun phrase, and is referred to using a plural pronoun. However, if the noun phrase is a measure noun phrase then it refers to an amount of 12 dollars; this could be made up of 12 dollar coins, 6 two dollar coins, a 10 dollar note and some change or even a check. In this case the amount must be referred to using a singular pronoun *it*.

These surprising properties support the need for a syntactic structure that is specific to this phenomenon. Ross (1995) showed that measure noun phrases are one of several classes of nouns that do not show the full range of behaviors that prototypical noun phrases do: they are defective noun phrases. In particular they do not exhibit all of the following behaviors: definite pronominalization; modifiability by a full range of determiners and modifiers; pluralizability and the triggering of number agreement; and the ability to undergo movement (such as passive, topicalization and various dislocations).

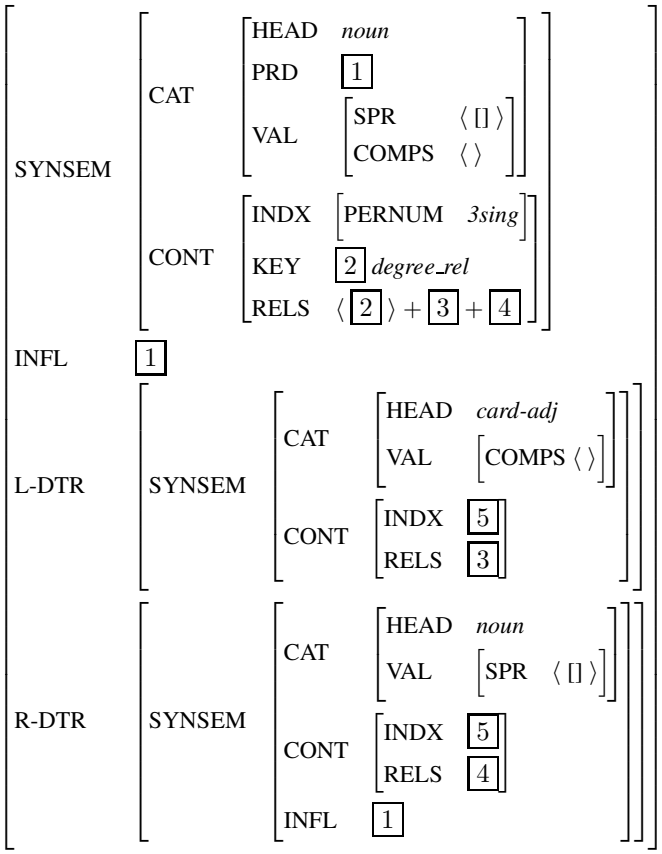
2 Analysis

We cannot introduce a syntactic rule that simply combines a number expression like *ten* or *two hundred fifty* with a count noun that agrees in number, where the rule's mother supplies the syntactic and semantic properties of a degree specifier. This would not allow for the number agreement mismatches: "*This ten feet is more frayed than the first piece of rope.*". Instead we need an exocentric rule that produces the singular N-bar *ten feet*, where the plurality of the right daughter noun is not preserved on the mother. We elaborate this rule slightly so the resulting measure phrase also has the necessary property to serve as a degree specifier: its distinguished semantic relation is a *degree_rel*.[1] The rule (2) builds a non-headed nominal phrase with two daughters, the left a numerical expression[2] and the right a noun.

[1]We adopt Minimal Recursion Semantics ((Copestake et al., 1999)) as the semantic framework for this analysis, but focus in this paper on the syntactic properties of measure phrases.

[2]We use the type *card-adj* (cardinal adjective) which includes numbers and other quantifiers such as *several*.

(2) Measure-Phrase Rule



This Measure Phrase (MP) rule admits the underlined phrases in (1) above, as well as in the following examples:

(3) a. *The building was several stories tall.*

b. *The first asteroid landed ten yards in front of me.*

c. *The next winning game would be just one city later.*

We note in passing that this binary-branching MP rule does not predict the grammaticality variation in examples where the measure noun appears alone, as in (4), but will focus in this paper only on an analysis of the binary structures.

(4) a. *The next village was miles away.*

b. * *The village was mile away.*

c. *He collapsed inches from the finish line.*

d. ? *He collapsed feet from the finish line.*

e. * *I need a cord that is feet long.*

f. * *Gallons is required to get me home.*

This MP rule must also account for the striking difference in inflection for measure phrases appearing in attributive vs. predicative adjective phrases, as illustrated in (5):

(5) a. *I bought a fourteen inch high candlestick.*
b. * *I bought a fourteen inches high candlestick.*
c. *I bought a candlestick that is fourteen inches high.*
d. * *I bought a candlestick that is fourteen inch high.*

While it might seem that the correlation between the attributive/predicative property and the form of the noun in the measure phrase involves singular vs. plural inflection on the noun, this would require a more complex interaction between the numerical adjective and the noun it modifies, even though this agreement follows just the expected pattern for the numeral *one* as seen in the contrast in (6).

(6) a. *I bought a candlestick that is one inch high.*
b. * *I bought a candlestick that is one inches high.*

If the *inch* in *fourteen inch high candlestick* is marked for singular number agreement, then the MP rule would have to require all numerals to appear with a singular noun if the measure phrase is in an attributive context, but to preserve the usual number agreement alternation when used in predicative phrases. Expressing such a feature co-occurrence restriction would require a complex distributed disjunction of values for the two features encoding number and predicativity, and would in our view lack explanatory force.

Instead, we argue that the nouns in measure phrases specifying attributive adjectives are in fact the stem forms, which are not yet inflected and hence underspecified for number. While most syntactic rules require that the daughter signs be fully inflected words or phrases, we propose that the MP rule is unusual in permitting an uninflected lexeme to enter the phrase structure without first undergoing inflectional morphology. Kiparsky (1982) suggests a similar approach where the left hand member in noun-noun compounds is uninflected. In particular, we assume that lexical signs bear the boolean (non-head) feature INFL, and that most entries in the lexicon are stem forms, marked [INFL −], with the inflectional rules producing from these stems syntactically admissible words which are [INFL +]. The MP rule simply requires that its nominal daughter identify its value for INFL with the mother's value for the boolean feature PRD used to distinguish attributive and predicative phrases, where this PRD feature is further identified for the two daughters in the general Specifier-Head rule which we use to combine *two foot* with *long* in *two foot long*. The left and right daughters do agree in number (INDX is co-indexed between them). However it is not co-indexed with the head (the head is always [INDX [PERNUM *3rdsing*]].

This analysis of the contrasts in (5) above in terms of inflection makes an interesting prediction about measure phrases containing nouns which ordinarily only

appear with plural inflection, such as *scissors*. Consider the examples in (7), which exhibit nearly the predicted judgments, where (7a) would be surprising if the nominal daughter in attributively used measure phrases had been constrained to be singular as in the candidate analysis we rejected:

(7) *Though most sewing cabinets have a small drawer large enough for only one pair of scissors,*

a. *my sister's cabinet has a three scissor wide drawer.*

b. ? *my sister's cabinet has a three scissors wide drawer.*

c. *my sister's cabinet drawer is three scissors wide.*

d. * *my sister's cabinet drawer is three scissor wide.*

Example (7b) appears to be acceptable to some speakers. This judgment can be accommodated in our proposed analysis by adding the form *scissors* to the lexicon as already marked for plural number but still keeping the strong assumption that all lexemes are constrained to be [INFL −] (analogous to Kiparsky's *mice*). Indeed, this measure phrase construction might serve as a useful source of illumination on the question of how lexically plural nouns ought to be defined within the lexicon.

We have proposed an analysis of measure phrases used as degree specifiers for adjectives, but have not yet accounted for the use of measure phrases as direct modifiers of nouns, as illustrated in (8):

(8) a. *I bought a fourteen inch candlestick.*

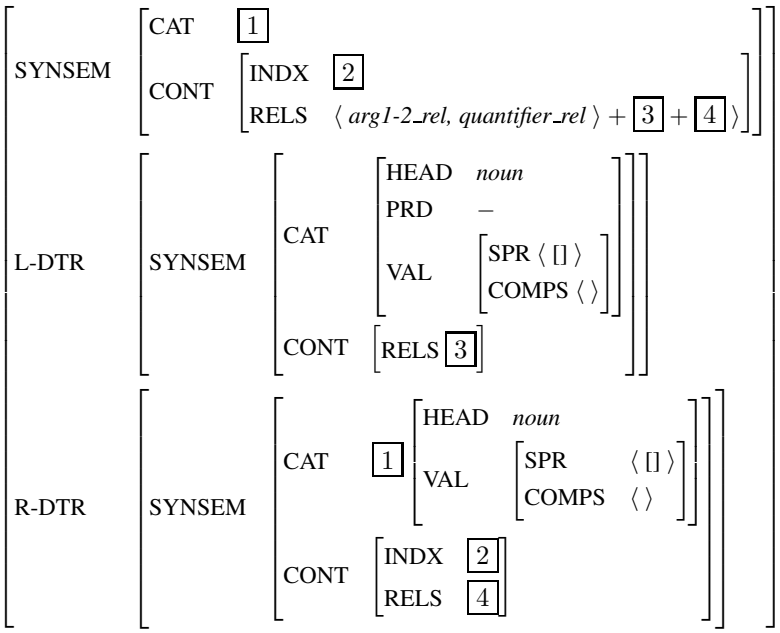
b. * *I bought a fourteen inches candlestick.*

c. *She lives in a six story building.*

d. * *She lives in a six stories building.*

We propose an analysis of noun-noun compound structures which constrains the left, non-head daughter to be attributive (in our notation, [PRD −]), sketched in (9). This, combined with the MP rule predicts exactly the data in (8). The measure phrase *fourteen inch* is perfectly suited to be the non-head daughter, while the [PRD +] phrase *fourteen inches* is not. Since the noun-noun compound rule already provides an underspecified two-place predicate (*arg1-2_rel*) relating the semantics of its two daughters, this same relation can also accommodate the underspecification of the dimension for which the measure phrase expresses a degree. That is, *a two foot cable* can describe either the length or the width of the cable (at least), and all we know from the compound construction is that whatever the dimension is, its extent is limited to two feet.

(9) Noun-noun Compound Rule



The noun-noun compound rule also provides a *quantifier_rel* to bind the instance variable of the non-head daughter.

Finally we account for the absence of attributive measure phrases as independent noun phrases, as shown in (10):

(10) a. *Fourteen inches is high enough.*

b. * *Fourteen inch is high enough.*

This contrast is obtained by constraining the MP rule so that the INFL value of the mother is the same as that of its nominal daughter. Therefore attributive (and hence non-inflected) measure phrases cannot be independent phrasal arguments, given our earlier assumption that most syntactic rules require fully inflected daughters.

Measure phrases (such as *twelve dollars* in *It costs twelve dollars*), in combination with their governing verb, predicate an amount; they do not refer to an entity or entities. Note, however, that a sentence like *I saw twelve dollars lying on the ground* has two interpretations. In one, the noun phrase is referential and there are twelve dollar notes or coins lying on the ground. In the other it is a measure phrase, in which case there could be 24 fifty cent coins, 12 dollar coins, 6 two dollar coins or any combination that adds up to a value of twelve dollars.

3 Implementation

A version of this analysis has been successfully implemented in a wide coverage grammar of English (ERG (Flickinger, 2000)) and tested extensively on data from

several domains, including meeting scheduling and travel planning (Verbmobil), and electronic commerce. For access to the implementation and the data, visit `http://lingo.stanford.edu`.

In the actual implementation, there are two measure-phrase rules which inherit from a common supertype `basic_measure_np_phrase`: the binary rule `binary_measure_np_phrase` which is presented above in (2), and a unary rule `bare_measure_np_phrase`.[3] The unary rule produces measure noun phrases with no explicit numerical phrase, as in (4) above and perhaps also in (11).

(11) *I read that years ago*

4 Distribution

We examined the distribution of the measure NP rule in a treebank (Oepen et al., 2002) of VerbMobil data (CD 6) (Wahlster, 2000). In 3,000 analyzed sentences there were 23 uses of the measure phrase rule in the top ranked parse (0.8%). A typical usage is given in (12).

(12) I attended *a two hour (long) meeting*.

The rule was also applied in at least one parse that was dis-preferred for 120 other utterances, such as (13).

(13) # *I attended a meeting later*.

5 Conclusions

We draw two conclusions from this analysis of measure phrases. First, we found evidence through this analysis that the constituents which appear in phrase structure are not always fully inflected, blurring the boundary between syntax and inflectional morphology. Second, it is clear that the surprising distributional characteristics of these phrases would be difficult to account for without positing an idiosyncratic syntactic rule of the kind we propose, lending support to the position that broad-coverage grammars will necessarily contain both highly schematic and highly idiosyncratic rules.

6 Further Work

We would like to extend this work in three ways. The first is to investigate the behavior of measure noun phrases in other languages. Preliminary investigations

[3] In written English, the noun can precede the number if it is a currency unit: *two hundred dollars* vs *$ 200*. This could be dealt with in two ways, either by using a preprocessor to rewrite the symbol into a word following the number, or to have two binary rules - the one presented in (2) and a second one, where the left and right daughters are reversed and the noun (right daughter) is constrained to be *common_noun_numcomp_synsem*, used for currency signs such as *$, £* and *AUD*.

into German show very different behavior: there is no link between inflection and predicativeness. Instead, there appears to be a sharp distinction between units and other nouns. Units (*meter*, ...) have no morphological singular/plural distinction (like English *sheep*) (14).[4] Other common nouns inflect for plural both in both predicative and attributive position (15). Semantically motivated discord is also present in German, although either singular or plural agreement is acceptable.

(14) a. *ein Fünfhundert-Euro-Artikel*
a five-hundred-Euro-article
A five hundred Euro article

b. *ein fünfhundert Euro teurer Artikel*
a five hundred Euro expensive article
An article costing five hundred Euros

c. *der Artikel kostest fünfhundert Euro*
the article costs five hundred Euro
An article costing five hundred Euros

d. *Fünfhundert Euro ist genug*
five hundred Euro is enough
Five hundred Euros is enough

e. *Fünfhundert Euro sind genug*
five hundred Euro are enough
Five hundred Euros are enough

f. *Fünfhundert Euros ist genug* (Euro coins)
five hundred Euros is enough
Five hundred Euros is enough

g. *Fünfhundert Euros sind genug* (Euro coins)
five hundred Euros are enough
Five hundred Euros are enough

(15) a. *ein Fünf-Seiten-Artikel*
a five-pages-article
A five hundred page article

b. *ein fünf Seiten länger Artikel*
a five pages long article
A five hundred page long article

c. *der Artikel ist fünf Seiten lang*
the article is five pages long
The article is five pages long

[4]The German word *Euro* has a plural, but it is only used when referring to the coins, not the currency.

The second extension is to investigate the history of this construction in English. It was suggested to us by Bill Ladusaw (p.c.) that it may be linked to the loss of case inflections over time.

Finally, we would like to examine the overall distribution of signs in corpora. The aim is to answer two questions: (1) How many rule-schema do we need? (2) How are they distributed? We hypothesize that the distribution follows a power law: a few constructions are used very frequently, and a great many are used infrequently. That is, we expect the distribution to be similar to that of words. Further, we expect an inverse correlation between how constrained a rule is and how frequent it is, although it is not yet clear how to quantify this. We will use the Redwoods Treebank data (`http://lingo.stanford.edu/redwoods/`) for the ERG as the basis for this line of investigation.

Acknowledgments

We would like to thank the members of the Lingo project, the audience at HPSG-2003, and Birte Lönneker for their helpful comments and discussion. This material is based upon work supported by the Research Collaboration between NTT Communication Science Laboratories, Nippon Telegraph and Telephone Corporation and CSLI, Stanford University.

References

Emily Bender and Dan Flickinger. Peripheral constructions and core phenomena. In Webelhuth et al. (1999), pages 199–214.

Francis Bond. *Determiners and Number in English contrasted with Japanese — as exemplified in Machine Translation*. PhD thesis, University of Queensland, Brisbane, Australia, 2001.

Ann Copestake, Dan Flickinger, Carl Pollard, and Ivan A. Sag. Minimal recursion semantics: An introduction. (manuscript `http://www-csli.stanford.edu/~aac/papers/newmrs.ps`), 1999.

Dan Flickinger. On building a more efficient grammar by exploiting types. *Natural Language Engineering*, 6(1):15–28, 2000. (Special Issue on Efficient Processing with HPSG).

Rodney D. Huddleston and Geoffrey K. Pullum. *The Cambridge Grammar of English*. Cambridge University Press, Cambridge, 2002. (principal authors).

Paul Kiparsky. Lexical phonology and morphology. In In-Seok Yang, editor, *Linguistics in the Morning Calm*, pages 3–91. Hansin, 1982.

Robert Malouf. *Mixed Categories in the Hierarchical Lexicon*. CSLI Publications, 2000.

Stephan Oepen, Dan Flickinger, Kristina Toutanova, and Christoper D. Manning. LinGO redwoods: A rich and dynamic treebank for HPSG. In *Proceedings of The First Workshop on Treebanks and Linguistic Theories (TLT2002)*, Sozopol, Bulgaria, 2002.

John P. Payne and Rodney D. Huddleston. Nouns and noun phrases. In *The Cambridge Grammar of English* Huddleston and Pullum (2002). (principal authors).

Carl Pollard and Ivan A. Sag. *Head Driven Phrase Structure Grammar*. University of Chicago Press, Chicago, 1994.

John Robert Ross. Defective noun phrases. In *Papers from the Regional Meeting of the Chicago Linguistic Society*, volume 31, pages 398–440. University of Chicago, 1995.

Ivan A. Sag. English relative clause constructions. *Journal of Linguistics*, 33(2): 431–484, 1997.

Jeffery D. Smith. English number names in HPSG. In Webelhuth et al. (1999), pages 145–160.

Wolfgang Wahlster, editor. *Verbmobil: Foundations of Speech-to-Speech Translation*. Springer, Berlin, Germany, 2000.

Gert Webelhuth, Jean-Pierre Koenig, and Andreas Kathol, editors. *Lexical and Constructional Aspects of Linguistic Explanation*. CSLI, Stanford, 1999.