Semantic construction for nominal expressions in cross-linguistic perspective

Emily Bender*, Markus Egg†, and Michael Tepper*

*University of Washington and †Universität des Saarlandes

1 Introduction

We discuss a case of typological variation which potentially constrains the design of the syntax-semantics interface: Turkish NPs with non-peripheral determiners, illustrated in (1).

(1) olası bir sonuç
probable one outcome
'one probable outcome'

On standard analyses (e.g., Barwise and Cooper (1981)), determiners introduce a (generalized) quantification, whose restriction is the semantics of the N plus any modifiers. After integrating the semantic contribution of the determiner, its restriction cannot be augmented further. Potential counterexamples to this prediction like (1) are thus touchstones for validity and generality of approaches to semantic construction. We sketch how Minimal Recursion Semantics (MRS) (Copestake et al., ip) and the Constraint Language for Lambda-Structures (CLLS) (Egg et al., 2001) handle the data, contributing to ongoing research on the comparison between these formalisms (Fuchss et al., 2004).

2 The MRS account

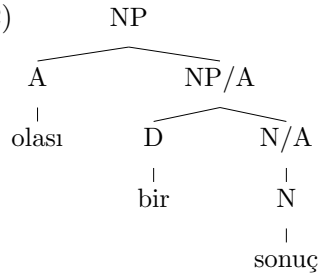
An MRS representation for a sign (word or phrase) consists of a list of elementary predications (RELS), each labeled with a *handle*; a set of scopal constraints among those handles (HCONS; called *qeq*s, for 'equal modulo quantifiers'); and a collection of information, called the *hook*, which is available for further semantic composition if the phrase is combined with other signs. The hook contains the semantic *index* and the highest scoping handle (other than quantifiers) of the sign.

In the MRS analysis of the English NP *a probable outcome*, the scopal adjective *probable* builds a qeq constraint between its argument position and the local top handle of the N *outcome*. The local top handle of *probable outcome* is then identified with the handle of *probable*, which in turn is related (by *a*) to the restriction of the quantifier by another qeq constraint. It would be impossible to construct the correct representation if the scopal adjective were to attach outside the determiner, as they appear to in Turkish.

If, however, the relative order of determiner and adjective in Turkish does not reflect tectogrammatical structure (i.e., the order of composition), there is no problem. One way to achieve this is with a sort of NP-internal extraction, giving trees like (2) for (1). The non-branching construction which takes N to N/A would in effect 'insert' the adjective's semantics in the

right place. Through the SLASH value, it places constraints on the HOOK and MOD values of the adjective, allowing the determiner to create the correct scopal relationship to the adjective.

(2)



If the word order in Turkish NPs were strictly Adj-Det-Noun, this would not be appealing. However, Turkish allows both orders of Adj and Det. As extraction tends to have pragmatic effects, we predict the Adj-Det order to be relatively marked. We tested this prediction with the METU Turkish treebank (Oflazer et al., 2003). In general, we find 65% Adj-Det-Noun. The most common determiner, *bir*, is a marker of indefiniteness and tends to appear after adjectives. All of the other determiners (except *hiçbir*, 'no', for which there are only two tokens), strongly prefer the opposite order:

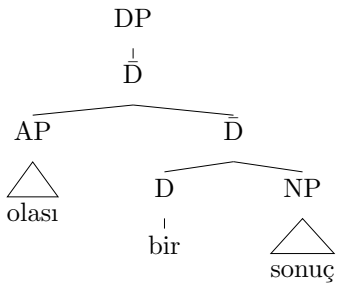
Det	Gloss	Det-Adj	N
kimi, şu	some, that	100%	3
bu	this	96%	50
o	that	83%	18
her	every	78%	9
hiçbir	no	50%	2
bir	a, one	21%	331

These data suggest that *bir* expresses only indefiniteness and not a quantifier. Note that Turkish has no overt definite determiner and that 88% of the 16,134 NPs in METU have no determiner. If the other determiners do contribute quantifiers, the prediction of the extraction-based account is borne out.

3 The CLLS account

CLLS (Egg et al., 2001) captures data like (1) by a flexible syntax-semantics interface. It derives underspecified semantic representations from a surface-oriented syntactic structure, e.g., (3) for (1). Here the AP modifies $\bar{\text{D}}$, but refers semantically only to the NP:

(3)



Such data are put down to *potential scope ambiguities*. Scope relations are encoded as *dominance relations*. They let non-quantifiers intervene scopally, thus, modeling scope by dominance (as opposed to qeq) relations is less constrained. This is crucial for the semantic construction for (1), which seems to contradict the claim that in practical applications these kinds of relations can be exchanged (Fuchss et al., 2004).

CLLS expressions are *constraints* on sets of semantic representations (*solutions*; here, λ-terms) with a *top* and a *bottom fragment*, e.g., the contributions of D and NP in the (simplified) constraint (4) for *bir sonuç*. The dominance relation (dotted line) in between allows intermediate scope of fragments added later in semantic construction:[1]

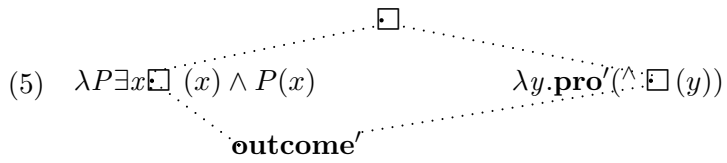
(4) $\lambda P \exists x.\ \boxdot\ (x) \wedge P(x)$

$\vdots$

$\mathbf{outcome}'$

The interface rule for modification then lets the modifier fragment dominate only the *bottom* fragment of the

[1] 'Holes' ($\Box$) are unknown parts of fragments. Fragments dominated by a hole are (im-)proper parts of what the hole stands for.



modified expression. E.g., (1) gets the semantic representation (5) above.

In prose: It is not yet clear what (5) stands for (there is only a hole on top) but the semantic contributions of D (right fragment) and AP (left fragment) are its immediate parts. The NP semantics (at the bottom) has narrowest scope. These structures emerge for quantifier scope ambiguities, too.

To derive the solutions of (5), dominance relations are strengthened to *identity*: Identifying D fragment and top hole, AP fragment and D fragment hole, and NP fragment and AP fragment hole, yields solution (6). Starting this procedure with the AP fragment is blocked by the types of the fragments involved: The hole in the AP fragment cannot be identified with the D fragment. This blocks spurious ambiguity.

(6) $\lambda P \exists x.\mathbf{pro}'(\hat{}\mathbf{outcome}'(x)) \wedge P(x)$

This approach extends directly to two closely related phenomena: Turkish relative-clause-like modifiers, e.g., gerunds, which precede determiners in nominal expressions (GER indicates the gerund suffix) and the modification of *indefinite pronouns* like *everyone*:

(7) a. yaz -dığ -ım her mektup
write GER my every letter
'every letter which I wrote'

b. everyone in this room

The modified DPs *her mektup* and *everyone* introduce a quantifier each, and the modifiers refer only to its *restriction*. The constraints for the modified expressions single out these restrictions in bottom fragments, then the modifiers intervene scopally between quantifier and restriction.

4 Conclusion

We have contrasted two approaches to semantic construction for Turkish nominal expressions and found that MRS is the more restrictive theory and CLLS the more flexible. Thus, MRS can make predictions about the distribution of the various word orders in the Turkish corpus, while CLLS is more straightforwardly applicable to related constructions. This result also highlights the importance of considering typologically diverse languages when designing semantic representations and syntax-semantics interfaces.

References

Barwise, J. and R. Cooper (1981). Generalized quantifiers and natural language. *Linguistics & Philosophy 4*, 159–219.

Copestake, A., D. Flickinger, C. Pollard, and I. Sag (i.p.). Minimal Recursion Semantics. An introduction. *Research on Language and Computation*.

Egg, M., A. Koller, and J. Niehren (2001). The constraint language for lambda-structures. *Journal of Logic, Language, and Information 10*, 457–485.

Fuchss, R., A. Koller, J. Niehren, and S. Thater (2004). Minimal Recursion Semantics as dominance constraints. In *Proceedings of ACL 2004*.

Oflazer, K., B. Say, D. Z. Hakkani-Tür, and G. Tür (2003). Building a Turkish treebank. In A. Abeillé (Ed.), *Treebanks: Building and Using Parsed Corpora*. Dordrecht: Kluwer.