Finding Depictive Secondary Predicates in Large Web Corpora

Benjamin Burkhardt[1], Laura Kallmeyer[1] & Timm Lichte[2]
([1]University of Düsseldorf, [2]University of Tübingen)

Introduction SECONDARY PREDICATES (SP), in the two flavors RESULTATIVE (RSP) and DEPICTIVE (DSP) as given in (1), have been discussed in the theoretical linguistics literature to some extent. Prototypical SPs in English are sentence final adjectival elements which modify one of the verb's nominal arguments:[1]

(1) a. Kim stomped the can_i $flat_i$.   (Resultative)
     b. Kim_i left the room $tired_i$.   (Depictive)

Concerning the distinction of DSPs and RSPs, semantically, RSPs specify the result state of a verb argument that is involved in the event denoted by the verbs, while DSPs specify properties of verb arguments that hold throughout the event of the verbal predicate but that are not directly connected to the event structure.

SP data in the literature is mostly introspectively constructed and relies on native speakers' grammaticality judgments. To our knowledge, a systematic corpus study of such phenomena has not been conducted so far. The goal of this paper is to provide such a study, investigating in particular whether the much-debated constructions in (2) can be observed:

(2) a. **Target Ambiguity:**
         Kim_i ate the $apple_j$ $unwashed_{i/j}$.
     b. **Depictive Stacking:**
         (i)   ? Kim_i ate the $steak_j$ raw_j $hungry_i$.
         (ii) ?? Kim_i ate the $steak_j$ $hungry_i$ raw_j.
         (iii) ?? Kim_i ate the $steak_j$ $salted_j$ raw_j.
     c. **Targeting of Unrealized Verbal Arguments:**
         The $game_j$ was played $barefoot_{i/*j}$.   (see Roberts 1987)
     d. **Targeting of Oblique Verbal Arguments:**
         (i)   Peter crashed into him_k $tired_{*k}$.
         (ii) The cash machine gave $John_k$ the money $hungry_{*k}$.
         (iii) You can't give $them_k$ injections $unconscious_k$.

While we already saw in (1) that SPs are not restricted to modifying only one verbal argument position, in (2a) we see an explicit example for a TARGET AMBIGUOUS DSP. In this case, without further context, it is unclear whether the adjectival element targets the subject or the direct object of the clause. Results of an informal pilot study with small group of American English native speakers, have shown that instances of DSP stacking as in (2b), have a lower acceptability value than simple cases of secondary predication. However they were not deemed unacceptable. Instead we observed different acceptability judgments given different linearizations of the subject and object targeting DSPs. In the case of (2b-i) with subject depictive following the object depictive, we saw higher acceptability values than with depictives occurring in the inverted order. Furthermore, in cases where two depictive elements target the same verb argument, we received the same decreased acceptability judgment values. Example (2c), which was originally discussed in Roberts (1987), shows that it is possible for DSPs to modify unrealized elements in the argument structure of the verb. This is based on a literal interpretation of the adjective in (2c), which would be semantically incompatible with the subject argument. Finally, when

---

[1]We will use co-indexation to mark the secondary predicate and its target.

confronted with examples of DSPs that seem to target the verb's oblique argument as in (2d-i), native speakers rejected these interpretations, even if only the oblique argument is compatible with the adjective as in (2d-ii). However, Simpson (2005) offers the counter example in the light verb construction clause given in (2d-iii).

A corpus study of the constructions discussed above comes with considerable difficulties due to the variability of the type of syntactic structures SPs can show. DSPs take the same position as adverbials (see (3)) and their linearization resembles the one of copula-like constructions (see (4)):

(3)   a.   Kim left$_i$ the room angrily$_i$.   (Adverbial)
       b.   Kim$_i$ left the room angry$_i$.   (Depictive)

(4)   a.   Kim got very drunk.   (Copula)
       b.   Kim$_i$ left very drunk$_i$.   (Depictive)

In order to meet these difficulties, we will make use of lemma information, POS tags and morphosyntactic annotation in our corpus filtering approach, operating on a large web corpus.

Method   As the basis of our study, we selected the ENCOW16AX, a large web corpus of English (see Schäfer 2015; Schäfer & Bildhauer 2012). The decision was driven by the following reasons: first, since the the corpus is web-based it covers a wide variety of Englishes. Secondly, the corpus covers the scale from formal to colloquial texts. Thirdly, since the considered SP constructions are rather rare, a large corpus was needed. ENCOW16AX comprises approximately 9,6 billion tokens. Fourth, since ENCOW16AX is dependency parsed using the MALT dependency parser and includes Penn Treebank POS tags and lemma information among others, it is well suited to filter out the target sentences that are of interest in this study. Additionally, the corpus is well documented, its creation pipeline is open source, and it is available for free download after registration.[2]

We conduct our corpus study by using a selection of 10 frequent adjectival stage level predicates such as *naked*, *hot*, *happy*, and *sick*. A preliminary analysis indicates that stage level predicates are more likely to appear in SP constructions. Even though the distinction between stage and individual level is context-dependent, the concept proves useful in our study.

The basic idea of trying to find SP constructions is to search for adjectival items in adverbial modifier positions which are not part of a copula construction. We applied three filter steps:

1. We identified all sentences which contained elements with the adjective POS tag which were also parsed to be heads in an adverbial modifier dependency. This relates to our observation about the syntactic position of SPs given above. At the same time, this step limits this study to SPs that consist of only one adjectival element.

2. Based on the output of the first step, three subcorpora were created. The first subcorpus includes sentences in which adjectival, adverbial modifier items are headed by verbal predicates tagged as heading a copula dependency relation. Sentences of this subcorpus mainly contain copula constructions, which are not of interest for this study. The second subcorpus includes sentences that contain copula tagged verbs as well as adjectival, adverbial modifier items independently from each other. The third subcorpus includes only sentences that contain adjectival, adverbial modifier items without any verbal predicates tagged as copula. In this subcorpus, the No-Copula subcorpus (NC), we expected to find the highest rate of SP constructions.

---

3. In the last filter step, we queried the NC subcorpus to receive samples of 200 sentences of a small number of hand selected highly frequent, argument oriented adjectives after adding some filter conditions based on the individual adjectives, for example to exclude multi-word expressions that contain the respective target adjective.

The filter steps given above involved checking of annotation information of the heads of target tokens, that is, when checking whether a copula or copula-like verb is the head of an adjectival adverbial modifier. At the same time, the filter steps also involved checking whether a target adjective token acts as the head of adverbial modifier relations in which the non-head was a copula or a copula-like verb. Since these kinds of queries proofed difficult to realize using the NoSketchEngine online interface of the ENCOW16A corpus, we developed custom filter scripts using the Python programming language.

Results & Implications In the first filter step we started with the ≈ 421 million sentences of the ENCOW16AX and found ≈ 4 million sentences containing at least one adjectival item that was part of adverbial modifier relation. Of these 4 million sentences ≈ 2.4 million were parsed as containing no copula constructions (NC subcorpus), which were used to take samples of 200 sentences per adjective. These samples were then annotated by hand for SP constructions. For the adjective *naked*, the sample yielded the following: 153 of the 200 sentences are SP constructions of which 144 are DSPs (6 of them fronted) and 9 are RSPs. The remaining 47 sentences included misidentified copula constructions, adnominal uses of the adjective, cases in which the adjective was used in a proper noun and small clauses. Some of the remaining cases were also excluded because they contained non-contemporary data, for example bible verses. Of the 144 DSP constructions, we found 119 which target the verb's ACTOR argument, 13 which target the UNDERGOER argument, and 4 in which an unrealized argument was targeted. Within this sample we found no examples of target stacking.

The findings for *naked* confirm the semantic analysis of Burkhardt, Lichte & Kallmeyer (2017), which proposes that a DSP can target either the actor or the undergoer, independent from the presence/absence of the respective argument in the syntactic structure. In the talk, we will report on the results for the selected adjectives and also discuss distributional differences between depictives which might be due to specific semantic constraints.

References • Burkhardt, Benjamin, Timm Lichte & Laura Kallmeyer. 2017. Depictives in English: An LTAG approach. In *Proceedings of the 13th International Workshop on Tree Adjoining Grammars and Related Formalisms*, 21–30. Umeå, Sweden: Association for Computational Linguistics. • Roberts, Ian. 1987. *The representation of implicit and dethematized subjects* (Linguistic Models 10). Berlin: De Gruyter Mouton. • Schäfer, Roland. 2015. Processing and querying large web corpora with the COW14 architecture. In Piotr Bański, Hanno Biber, Evelyn Breiteneder, Marc Kupietz, Harald Lüngen & Andreas Witt (eds.), *Proceedings of Challenges in the Management of Large Corpora 3 (CMLC-3)*. Lancaster: IDS. • Schäfer, Roland & Felix Bildhauer. 2012. Building large corpora from the web using a new efficient tool chain. In Nicoletta Calzolari (Conference Chair), Khalid Choukri, Thierry Declerck, Mehmet Uğur Doğan, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk & Stelios Piperidis (eds.), *Proceedings of the eight International Conference on Language Resources and Evaluation (LREC'12)*, 486–493. Istanbul, Turkey: European Language Resources Association (ELRA). • Simpson, Jane. 2005. Depictives in English and Warlpiri. In Nikolaus P. Himmelmann & Eva Schultze-Berndt (eds.), *Secondary predication and adverbial modification: The typology of depictives*, 69–106. Oxford: Oxford University Press.