The syntax-morphology divide is gradient and sensitive to frequency: Evidence from suspended affixation

Eva Neu[1], Utku Turk[2], Ozge Bakay[1], Brian Dillon[1], Gaja Jarosz[1]

[1]*University of Massachusetts Amherst,* [2]*University of Maryland, College Park*

Introduction. Suspended affixation (SA) is a phenomenon at the boundary between syntax and morphology where coordination – a syntactic operation – conjoins parts of words. In Turkish, SA is highly productive with inflectional morphemes, but idiosyncratically restricted with derivational morphemes. This effect is incompatible with the view that derivational morphology is a separate module completely inaccessible to syntax, but also with the view that syntax and morphology form a single module in which coordination can operate freely. Here we explore the idea that the divide between syntax and morphology is gradient and sensitive to frequency. We provide experimental evidence that more frequent coordinations are more acceptable under derivational SA and connect this finding to previous work on degrees of lexicalization.

Background I: Suspended affixation. In SA, an affix or string of affixes takes scope over a conjoined phrase. In (1), the plural morpheme can be interpreted on both conjuncts. SA has variably been analyzed as base-generated (Dolatian, 2022; Kabak, 2007), with the affix merged in a position c-commanding both conjuncts, or as ellipsis (Despić, 2017; Erschler, 2018; Guseva & Weisser, 2018), with the

(1)   [kedi ve köpek]-*ler*
       cat   and dog-PL
       'cat and dogs'/'cats and dogs'

affix on the first conjunct eliding under identity with the second. Under either analysis, SA raises challenges for a strict modular distinction between syntax and morphology that predicts words to be inaccessible to syntactic processes such as coordination and ellipsis (Akkuş, 2016).

Kornfilt (2012) has argued that Turkish SA is only licensed with inflectional, not derivational affixes (2).

(2)   *[don-dur-up   kızar-t]-*ma*
       freeze-CAUS-and roast-CAUS-RESULT
       intended: 'ice cream and roast meat'

She concludes that only inflectional morphology takes place in the syntax, while derivational morphology is a separate module inaccessible to syntactic operations like coordination and ellipsis (see also Anderson, 1992). However,

derivational SA is licensed in isolated examples (3). Kornfilt argues that these are exceptions stored idiomatically in the lexicon. But Akkuş provides a wide range of examples sourced online such as (4) and notes that they are both more common than previously assumed and often remain acceptable with minor changes. This suggests that derivational SA is not simply reducible to rigid, idiomatic exceptions.

(3)   [ana   ve baba]-*lık*
       mother and father-DER
       'mother- and fatherhood'

(4)   [arkadaş ve dost]-*ça*   bir hava
       friend   and buddy-DER an atmosphere
       'a friendly and intimate environment'

Akkuş takes these findings as evidence against a strict syntax-morphology divide and for single-module architectures like Distributed Morphology (Halle and Marantz, 1993, 1994). However, compared to inflectional SA, derivational SA is still extremely restricted. In particular, Akkuş's examples suggest that it is licensed predominantly with high-frequency coordinations. If syntax can operate freely on any morpheme, this effect of lexical specificity for derivational morphemes only remains unaccounted for.

Background II: Degrees of lexicalization. To solve this puzzle, we draw a connection to work by Morgan & Levy (2015, 2016), who show that more frequent coordinated phrases have more rigid ordering preferences. E.g., *ladies and gentlemen* is highly likely to occur with this particular order, whereas the order of *towels and sunglasses/sunglasses and towels* is more fluid. Morgan & Levy take this as evidence for a trade-off between compositional and stored, holistic representations of complex phrases: less frequent expressions are more likely to be decomposed, and more frequent expressions to be represented holistically, resulting in increasingly frozen ordering preferences. We propose that varying degrees of lexicalization can account for the limited availability of derivational SA. Derivational affixes sits closer to the root than inflectional affixes (Arad, 2003), indicating that they combine with a highly lexicalized constituent with a strongly entrenched holistic representation. Assuming a base generation analysis of SA, derivational affixes are thus expected to only be able to merge with a larger coordinated phrase such as *mother and father* or *friend and buddy* if

this coordinated phrase is sufficiently frequent and lexicalized. Therefore, we predict that more lexicalized binomials should be more acceptable under derivational SA. Next, we test this prediction experimentally.

Methods I: Corpus data. Following Morgan & Levy, we operationalize degree of lexicalization as the frequency of a binomial expression. We collected the frequencies of 32 binomials from the tagged Turkish Web 2012 corpus ($\approx$ 3.4 billion words). Specifically, for each pair of nouns, we count how often the two stems, whether or not bearing any affixes, appear in the sequence 'STEM.1 and STEM.2.' The frequencies we collect range from 131657 (*mother and father*) to 1 (*mouse and lion*). We ensure that even low-frequency coordinations are semantically natural, with the nouns being taken from the same conceptual realms.

Methods II: Judgment Experiment. We compared the acceptability of the 32 binomials chosen above with SA vs. without SA. To control for suffix type, we used four different highly productive derivational suffixes that formed adjectives: *-cI* (supporter), *-lI* (with), *-sIz* (without) and *-lIk* (hood). 47 Turkish speakers rated 32 experimental and 64 filler sentences, each with a distinct binomial, in a 4x2 within-subject Latin square design and 64 fillers on a 6-point Likert scale (6='very good'). Examples for *-lI* and *-sIz* are given in (5).

(5)   Selin   tamir   için   [kapı(-*lı/sız*)   ve   pencere]-*li/siz*   bir   atölye   hazırlamış.
       Selin   repair   for   door(-*having/-less*)   and   window-*having/-less*   an   atelier   prepared
       'Selin prepared a [door(-*having/-less*) and window]-*having/-less* atelier for the repair.'

Analysis and results. Our results (Fig1a) showed that SA generally reduced the acceptability for *-sIz* and *-cI*, but not for *-lI* or *-lIk*. Crucially, for *-sIz* and *-cI*, the effect of frequency was not uniform: while acceptability of SA increased with frequency, in line with predictions, there were negligible effects of frequency on sentences without SA. These results were verified by our Bayesian cumulative link models, with rank frequency as a predictor for each SUFFIX (see Fig1b). Frequency was entered as a centered rank frequency (lower rank equals more frequent) measure based on binomial counts in the corpus. Our model verified these findings. We found a strong positive interaction between frequency and SA in both *-sIz* ($\hat{\beta}$=0.06, $P(\hat{\beta}>0)$=0.98) and *-cI* ($\hat{\beta}$=0.04, $P(\hat{\beta}>0)$=0.88), but not such interaction for *-lIk* ($\hat{\beta}=-0.01$, $P(\hat{\beta}>0)$=0.31) and reversed interaction for *-lI* ($\hat{\beta}=-0.05$, $P(\hat{\beta}\leq0)$=0.93). We also verified the effect of suspended affixation only for *-sIz* ($P(\hat{\beta}>0)>0.99$) and *-cI* ($P(\hat{\beta}>0)>0.99$).
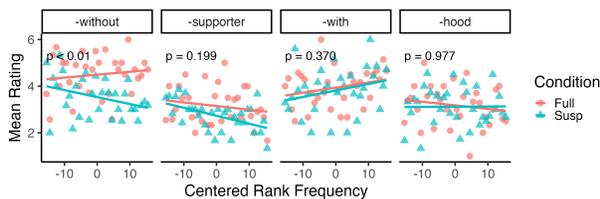
(a) Mean ratings for binomials with and without SA, shown separately for each suffix. Items are ordered from least (left) to most frequent binomial (right). Fitted lines indicate linear trends in ratings.
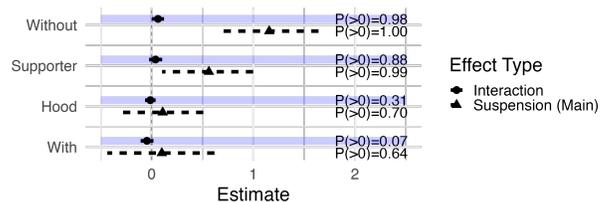
(b) Posterior coefficients estimates for the main effect of Suspension and Interaction with Frequency. Interactions and probability of a positive interaction is highlighted.

Discussion. The results support our hypothesis that derivational SA is more acceptable with high frequency binomials, although this effect was limited to certain suffixes. We argue that the reason why SA with *-lI* and *-lIk* was not susceptible to frequency is that with these affixes, SA hardly affects acceptability in the first place (*contra* previous work). Thus, these affixes may behave like inflectional morphemes, which generally participate in SA freely and do not seem sensitive to the frequency of the coordinated phase. Importantly, our prediction of frequency effects presupposed an *available* base generation analysis, under which the coordinated phrase without the affix forms a constituent that can be represented more or less holistically. An ellipsis analysis predicts no such frequency effects of the coordination. This suggests that *-sIz* and *-cI* can participate in SA in a base generation structure, while inflectional and some other derivational affixes such as *-lI* and *-lIk* also allow for an ellipsis structure. Why this should be the case is a question we leave for future work. We also plan to test frequency effects for a wider range of affixes, including inflectional morphemes.