**SPLIT NP TOPICALIZATION IN VIETNAMESE**

Keywords: Vietnamese, split NP, Copy Deletion, doubling

**Obligatory doubling/deletion.** Topicalization of N from a numeral (num) + classifier (K) + NP complex in Vietnamese shows an intricate pattern of pronunciation. If N is relational and has a complement, N must be pronounced at both the derived and the base position (1). If N is non-relational, such doubling is impossible (2).

(1) $[_{CP}$ bạn $[_{\overline{C}}$ thì $[_{TP}$ nó sẽ gặp $[_{NumP}$ hai $[_{KP}$ người $[_{NP}$ *(bạn) $[_{PP}$ của John ]]]]]]]
friend TOP he will meet two K friend of John

(2) $[_{CP}$ mũ $[_{\overline{C}}$ thì $[_{TP}$ nó sẽ mua $[_{NumP}$ hai $[_{KP}$ cái $[_{NP}$ (*mũ) ]]]]]]
hat TOP he will buy two K hat

We propose to derive the facts in (1) and (2) from the following two claims about Copy Deletion, the operation which eliminates the lower copy of a chain from the phonological representation (cf. Chomsky 1995).

(3) Right Edge Condition (REC): Copy Deletion can apply to a chain $(\alpha, \beta)$ only if the lower copy $\beta$ ends an XP, i.e. only if the rightmost morpheme of $\beta$ coincides with the rightmost morpheme of a non-projecting category.

(4) Minimize Pronunciation (MP): Copy Deletion must apply when it can.

Because the lower occurence of *bạn* 'friend' in (1) does not end an XP, Copy Deletion cannot apply, and doubling results. The lower occurence of *mũ* 'hat' in (2), on the other hand, is an XP itself, hence does end an XP in the sense of (3). Consequently, Copy Deletion can, and by virtue of MP must, apply in (2).

**Optional doubling/deletion.** Other facts about split topicalization in Vietnamese follow from REC and MP. Thus, doubling of the topicalized N is optional if N is a non-relational noun followed by a modifier (5).

(5) mũ thì nó sẽ mua hai cái (mũ) màu đỏ
hat TOP he will buy two K hat red

We assume that nouns in Vietnamese, a classifier language, denote cumulative predicates whose extension include both singular and plural individuals, and that the classifier's function is to map a cumulative predicate to an atomic one whose extension includes only singular individual (Chierchia 1998). Thus, $[\![mũ]\!]$ = [$\lambda$x. x is a singular hat or a plurality of hats] and $[\![K]\!]$ = [$\lambda$P. $\lambda$x. P(x) and x is singular]. This means that both (6a) and (6b) are interpretable structures denoting the set of singular red hats, as both N and [K N] are of type <e,t> and hence can compose with the AP via Predicate Modification (Heim & Kratzer 1998).

(6) a. $[_{KP<e,t>}$ $[_{KP<e,t>}$ $K_{<et,et>}$ $hat_{<e,t>}]$ $[_{AP<e,t>}$ red]]
b. $[_{KP<e,t>}$ $K_{<et,et>}$ $[_{NP}$ $hat_{<e,t>}$ $[_{AP<e,t>}$ red]]]

Fronting N from (6a) triggers Copy Deletion as N is at the right edge of NP. Fronting N from (6b) inhibits Copy Deletion as N is not at the right edge of any XP. Thus, optionality of doubling in (5) is the consequence of KP in (5) being analyzable as either (6a) or (6b).

Optional doubling is also attested when the topicalized N is a relational noun without an overt complement (7).

(7) bạn thì nó sẽ gặp hai người (bạn)
friend TOP he will meet two K friend

Assuming *bạn* 'friend' denotes the relation [$\lambda$x. $\lambda$y. y is friends with x] and is hence of type <e,et>, (7) is only interpretable if KP has the parse in (8), where *bạn* merges with a null pronoun, *pro*, which is of type e. (We assume that *pro* is existentially quantified over within the NP, so the NP in (8) would denote the predicate [$\lambda$x. $\exists$y. x is friends with y], for example.)

(8) $[_{KP<e,t>}$ $K_{<et,et>}$ $[_{NP<e,t>}$ $friend_{<e,et>}$ $pro_{e}$ ]]

The non-doubling option of (7) corresponds to the option of fronting the whole NP from (8). The doubling option corresponds to the option of fronting just N, leaving *pro* behind.

**Discourse coherence.** The sentences in (9) are perceived as felicitous discourse while (10) is not (the subscripts "F" indicates the focus of the sentence.)

(9) a. Vợ thì nó gặp $hai_F$ người. Bạn thì nó gặp $ba_F$ người.
wife TOP he met two K friend TOP he met three K
b. Vợ thì nó gặp $hai_F$ người vợ. Bạn thì nó gặp $ba_F$ người bạn.
wife TOP he met two K wife friend TOP he met three K friend

(10) #Vợ thì nó gặp $hai_F$ người. Bạn thì nó gặp $ba_F$ người bạn.
wife TOP he met two K friend TOP he met three K friend

It follows from what has been said that (9a) and (9b) must be parsed as (11a) and (11b), respectively, while (10) must be parsed as (12), which consists of the first sentence of (11a) and the second sentence of (11b), in that order. (The application of Copy Deletion is represented by ~~strikethrough~~.)

(11) a. $[_{S_1}$ wife *pro* TOP he met $two_F$ K ~~$[_{NP}$ wife *pro*~~]]. $[_{S_2}$ friend *pro* TOP he met $three_F$ K ~~$[_{NP}$ friend *pro*~~]].
b. $[_{S_3}$ wife TOP he met $two_F$ K $[_{NP}$ wife *pro*]]. $[_{S_4}$ friend TOP he met $three_F$ K $[_{NP}$ friend *pro*]].

(12) [$_{S_1}$ wife *pro* TOP he met two$_F$ K ~~[$_{NP}$ wife *pro*]~~]. [$_{S_4}$ friend TOP he met three$_F$ K [$_{NP}$ friend *pro*]].

Taking this result as starting point, we propose to account for the contrast between (9a-b) and (10) in terms of Büring's (1999, 2003) theory of discourse structure. For lack of space, we cannot go into the details of Büring's theory here. Suffice it to say that this theory can be applied to the data at hand in such a way as to make the following predictions ($S_1$, $S_2$, $S_3$ and $S_4$ refer to the sentences in (11) and (12)).

(13) a. A sequence of sentences <$S_i$, $S_j$>, both of which contain a contrastive topic, is only felicitous if $S_i$ and $S_j$ have the same topic value, i.e. if $[\![S_i]\!]^{top} = [\![S_j]\!]^{top}$

b. $[\![S_1]\!]^{top} = [\![S_2]\!]^{top}$ = {how many P did he meet? | P is of type <e,t>}

c. $[\![S_3]\!]^{top} = [\![S_4]\!]^{top}$ = {how many R *pro* did he meet? | R is of type <e,et>}

It follows from (13) that both <$S_1$, $S_2$> and <$S_3$, $S_4$> are felicitous while <$S_1$, $S_4$> is not, which is what we observe in (9) and (10).

**Ellipsis.** A potential counterexample to our proposal is the felicity of (14). However, we claim that (14) involves KP-ellipsis and is to be analyzed as (15).

(14) Vo thi no gap hai$_F$ nguoi vo. Ban thi no gap ba$_F$.
wife TOP he met two K wife friend TOP he met three

(15) wife TOP he met two$_F$ [$_{KP}$ K [$_{NP}$ wife *pro*]]. friend TOP he met three$_F$ ~~[$_{KP}$ K [$_{NP}$ friend *pro*]]~~

It is then predicted that (16a) is degraded because it involves cataphoric ellipsis (cf. **He read three. I read two books*) but (16b) is better than (16a) as cataphoric ellipsis improves under embedding (cf. *He read three but I read only two books*). Both predictions are born out.

(16) a. #Bạn thì nó gặp ba$_F$. Vợ thì nó gặp hai$_F$ người vợ.
friend TOP he met three wife TOP he met two CL wife

b. Bạn thì nó gặp ba$_F$ nhưng vợ thì nó chỉ gặp hai$_F$ người vợ.
friend TOP he met three but wife TOP he only met two CL wife

**Measure words.** Container nouns such as *thùng* 'box' can be used as "measure words," i.e. can take on a meaning of type <et,et> akin to that of a classifier (17).

(17) $[\![$*thùng*$]\!]$ = [$\lambda$P. $\lambda$x. x is a box-load of P]

Thus, *hai thùng sách* 'two box book' would mean 'two box-loads of books.' Topicalization of N out of a numeral + measure word + NP complex shows a puzzling correlation between pronunciation and interpretation: while the doubling pattern (18a) is ambiguous, the non-doubling pattern (18b) is not.

(18) a. sách thì nó sẽ mua hai thùng sách to
book TOP he will buy two box book large
'he will buy two large box-loads of books' / 'he will buy two box-loads of large books'

b. sách thì nó sẽ mua hai thùng to
book TOP he will buy two box large
'he will buy two large box-loads of books' / *'he will buy two box-loads of large books'

To account for this observation, we assume that when a measure word M merges with a nominal, either can project, so that we have the following four analyses for *hai thùng sách to* 'two box book large.'

(19) a. [$_{Num}$ two [$_M$ box [$_N$ book large]]]
b. [$_{Num}$ two [$_N$ box [$_N$ book large]]]
c. [$_{Num}$ two [$_M$ [$_M$ box book] large]]
d. [$_{Num}$ two [$_N$ [$_N$ box book] large]]

Given the meaning of *thùng* 'box', $[\![$(19a)$]\!]$ = $[\![$(19b)$]\!]$ = 'two box-loads of large books' and $[\![$(19c)$]\!]$ = $[\![$(19d)$]\!]$ = 'two large box-loads of books.' Fronting *book* from (19a), (19b) or (19d) will result in doubling since the lower copy does not end a non-projecting category. This is why the doubling pattern is ambiguous. Fronting *book* from (19c) results in deletion of the lower copy, as this copy is itself a non-projecting category and thus ends one in the sense of (3). This is why the non-doubling pattern is unambiguous and has the meaning that it does.

**Conclusion.** We account for a series of sound-meaning correlations observed in NP split topicalization constructions in Vietnamese. Our explanation relies crucially on the idea that movement is "internal merge," resulting in one constituent occupying two positions in the phrase marker (cf. Chomsky 2013), and on the idea that Copy Deletion, presumably a syntax-phonology mapping rule, makes reference to projection levels and the linear order of constituents (cf. Selkirk 1986, Selkirk & Tateishi 1991). To the extent that our proposal is correct, the set of data from Vietnamese that we examine here is to be considered evidence for these ideas.

**Büring, D. (1999)** *Topic*. **Büring, D. (1999)** *On D-Trees, Beans and B-Accents*. **Chierchia, G (1995)** *Reference to Kinds Across Languages*. **Chomsky, N (1995)** *The Minimalist Program*. **Chomsky, N. (2013)** *Problems of Projection*. **Heim, I. & A. Kratzer (1998)** *Semantics in Generative Grammar*. **Selkirk, L. (1986)** *On derived domains in sentence phonology*. **Selkirk, L. & Tateishi, K. (1991).** *Syntax and downstep in Japanese*.