Verb phrase ellipsis avoids troughs in the ID profile: An information-theoretic account to
VPE based on evidence from rating and reading time data

Lisa Schäfer, Robin Lemke, Heiner Drenhaus and Ingo Reich (Saarland University)

lisa.schaefer@uni-saarland.de

Background. While the grammar of verb phrase ellipsis (VPE) is well-studied from a theoretical
and a psycholinguistic perspective (see Reich (2011); Phillips and Parker (2014) for an overview),
this is not true for the question of why a speaker uses an ellipsis at all. However, when contrasting
the full form (1a) with the corresponding VPE (1b) an obvious hypothesis is that the speaker follows
a pragmatic imperative for rational communication: Avoid redundancy!

(1) a. John played football in the backyard and Bill played football in the backyard, too.
b. John played football in the backyard and Bill did, too.

Account. This strategy may be modeled by the information-theoretic concept of Uniform Informa-
tion Density (UID) (Levy and Jaeger, 2007). According to UID, speakers tend towards distributing
information as uniformly as possible across utterances avoiding troughs in the information density
(ID) profile (Fig. 1), i.e. regions with continuously low information. Troughs are caused by redun-
dant information like the second conjunct of (1a) and are the longer the longer the redundant part
is. We conducted an acceptability rating and a self-paced reading task that allow us to correlate
preferences for VPE with cognitive effort indexed by reading times.

Exp. 1. We conducted the rating study in a 2 x 2 (FORM: ellipsis vs. full form x LENGTH: short (no
PP) vs. long (with PP)) within-subjects design. In the short condition we presented a plain VP (2a
& 2c), that was expanded by a PP in the long condition (2b & 2d). We expect a relatively higher
preference for VPE in the long condition due to the longer trough in the ID profile. 41 British English
native speakers recruited via Prolific Academic (Palan and Schitter, 2018) rated 32 items like (2)
and 72 fillers (i.a. coordinated structures with two distinct VPs (3)) on a 7-point Likert scale (7 =
completely natural). We analyzed our data with CLMMs (ordinal, Christensen (2018)) in R with a
full random effects structure (Barr et al., 2013) and compared models with likelihood ratio tests: A
significant main effect of LENGTH ($\chi^2 = 27.92$, p < .001) shows that overall items with longer VPs
were rated worse (Fig. 2). A significant FORM:LENGTH interaction ($\chi^2 = 7.66$, p < .01) indicates that
the full form was rated particularly worse in the long condition. This is in line with our hypothesis:
Speakers prefer VPE the more the more redundant the repetition of the antecedent VP would be.

Exp. 2. Our account predicts that this preference is caused by the intention to avoid troughs in
the ID profile. We used a self-paced reading study on the full forms (2a vs. 2b) to test whether the
redundant VP indeed creates such a trough in the second conjunct. We expected this trough to
be larger in the more redundant long condition, which would be indicated by a larger difference
in average reading time between the first and the second conjunct in the long condition. Using
IBEX (Drummond, 2016), the items and fillers from exp. 1, extended by a spillover region, were
presented word-by-word and centered to 87 native speakers of British English recruited via Prolific
Academic. Our dependent variable were the cumulated reading times per CONJUNCT (italicized
in (2)) which we residualized per subject (Gibson and Levy, 2016). We analyzed the data with
linear mixed effect models (lme4, Bates et al. (2015)) in R with random intercepts for subjects and
items and compared models with likelihood ratio tests. A significant main effect of CONJUNCT ($\chi^2 =
159.18$, p < .001) (Fig. 3) shows that the second conjunct was generally read faster than the first. A
significant interaction between LENGTH:CONJUNCT ($\chi^2 = 63.04$, p < .001) indicates that the second
conjunct was especially faster in the long condition. This supports our hypothesis: The longer VP
is more redundant and, thus, creates a longer trough, which is reflected in faster reading times.

Discussion. Our data provide further evidence for the effect of UID on encoding preferences:
Speakers omit redundant parts of the utterance and use ellipsis to avoid troughs in the ID profile.
This preference for omission increases the larger the redundant part is.
(2)  
a. John played football and Bill played football, too. (full form, short (no PP))
  b. John played football in the backyard of the house and Bill played football in the backyard of the house, too. (full form, long (with PP))
  c. John played football and Bill did, too. (ellipsis, short (no PP))
  d. John played football in the backyard of the house and Bill did, too. (ellipsis, long (with PP))

(3)  
Ann packed provisions on the night before the trip and Tom loaded the car on the night before the trip (since they wanted to depart early in the morning). (filler (with spillover))