

Referential Form, Word Order, and Implicit Causality in Turkish Emotion Verbs

Duygu Sarisoy (duyguo@metu.edu.tr)

Department of Foreign Language Education, Middle East Technical University
Dumlupinar Blv, 1/6 D:133, Ankara 06800, Turkey

Ebru Evcen (eevcen@ucsd.edu)

Department of Linguistics, UC San Diego
9500 Gilman Drive, San Diego, CA 92093 USA

Joshua Hartshorne (Joshua.hartshorne@hey.com)

MGH Institute of Health Professions, Communication Sciences and Disorders
36 1st Ave, Boston, MA 02129 USA

Abstract

Pronouns are unique in discourse, as their meaning depends almost entirely on context. Early theories provided simple accounts of how meaning is determined, but research has revealed complex influences across syntax, semantics, discourse, and pragmatics. Evaluating theories is challenging due to methodological inconsistencies and a focus on English, limiting generalizability. Here, we take a step towards a clear empirical foundation for theory, with a tightly controlled study of comprehension of overt and null pronouns in Turkish. We show that pronoun resolution in Turkish is influenced by verb type, word order, and referential form, though not always in ways predicted by existing theories. Our findings highlight the need for further cross-linguistic research to refine models of pronoun interpretation and better account for the interaction of syntactic and discourse factors.

Keywords: implicit causality; anaphora resolution; pronoun interpretation; Turkish emotion verbs

Introduction

Third-person pronouns (he/she/they/it), despite their ubiquity in discourse, are a profound challenge to the language sciences. Formally ambiguous, their use and interpretation depend almost entirely on context. Influential early accounts told a fairly simple story: pronouns are preferentially used to refer to the topic (or, variably, the focus, center, or most prominent entity) of the previous clause (in English, this will typically be the grammatical subject) and listeners will similarly interpret pronouns in a complementary manner (Ariel, 1990; Givón, 1976; Gundel et al., 1993; Grosz et al., 1983). Conversely, full noun phrases are used to refer to non-topics (or non-foci, etc.), and listeners are aware of this fact, perhaps through applying Gricean principles (“If the topic was the same, the speaker would have used a pronoun...”; Gundel et al., 1993).

Despite some initial empirical successes, such theories quickly ran up against significant challenges. First, while it is

true that in certain types of discourses — especially those in newspaper articles, the subject of much of the initial quantitative research — successive sentences tend to have the same topic, it is emphatically not true in general. Arnold (1998) analyzed a corpus of children’s books, finding 64% of pronouns referred to the previous subject — a majority, certainly, but hardly overwhelming.

This appears to be partly a function of discourse structure. Kehler (2002) catalogs over a dozen different kinds of discourse structures: some, such as an “elaboration”, consist of a set of statements about the same thing and thus naturally involve repeated topics (*Jeremiah was a bullfrog. He was a good friend of mine.*). Others (such as “explanation”) have other purposes (like providing the explanation of an event) for which topic-preservation may be counter-productive (e.g., *The feather broke the vase. It was incredibly destructive?*). Empirically, manipulating the discourse structure has dramatic effects on pronoun interpretation (Crinean & Garnham, 2006; Hartshorne et al., 2015; Kehler, 2002; Kehler et al., 2008; Stevenson et al., 1994). Compare, for instance *Al frightened Bart because he drank all the wine* with *Al frightened Bart, so he drank all the wine*. Most readers will conclude that Al drank all the wine in the first sentence, but that Bart did in the second.

Second, even considering the kinds of sentences the earlier theories were developed for (elaboration discourses), theorists remain divided as to the relevant construct (topic, focus, center, etc.) and how it should be defined. In English, many of these constructs are confounded: the subject is usually the topic and it is also the first-mentioned entity. Studies of other languages where these things dissociate have suggested that each plays a distinct role in pronoun use and comprehension (Crawley & Stevenson, 1990; Gordon & Chan, 1995; Järvikivi et al. 2005; Foraker & McElree, 2007; Kaiser & Trueswell 2008; Walker et al., 1994).

Third, even controlling for discourse structure, the meaning of the sentences can dramatically change pronoun

interpretation (Au, 1986; Crinean & Garnham, 2006; Ferstl et al., 2011; Garnham et al., 2021; Garvey & Caramazza, 1974; Hartshorne & Snedeker, 2013; Hartshorne et al., 2013; Hartshorne et al., 2015; Stevenson et al., 1994). For instance, *Al frightened Bart because he drank all the wine* with *Al feared Bart because he drank all the wine*. In explanation discourses like these, changing the verb in the first clause often (not always) changes interpretation of pronouns in the second, an effect called “implicit causality” (Garvey & Caramazza, 1974). To complicate matters, the same manipulation of verbs can have the opposite effect in other discourse contexts, or none at all (Au, 1986; Crinean & Garnham, 2006; Hartshorne et al., 2015; Stevenson et al., 1994). These effects are sometimes given different names for different discourse contexts (the effect of Clause 1 verb on Clause 2 pronoun in a result discourse like *Al frightened Bart, so he drank all the wine* is sometimes called “implicit consequentiality”; Crinean & Garnham, 2006), however many remain unnamed. Here, we will use the more generic term “verb bias” to refer to all such effects.¹

Some initial theoretical accounts attributed these verb biases to thematic role (Brown & Fish, 1983). For instance, perhaps pronouns in explanation discourses preferentially refer to the AGENT of an AGENT-PATIENT verb. However, such accounts were always conjectural and sometimes circular (a frequently-invoked thematic role, EVOCATOR, is defined in terms of its effect on pronoun comprehension and does not otherwise appear in linguistic theory; see Hartshorne & Snedeker, 2013, for discussion). The only studies to systematically investigate the role of thematic role found that thematic role accounts predict the data at or below chance levels (Hartshorne & Snedeker, 2013; Hartshorne et al., 2015). In contrast, Hartshorne and colleagues found that --- at least for explanation and result discourses --- verb biases are tightly linked to the verb’s Levin argument structure class. This is consistent with a general move away from thematic roles in linguistic theory, as such theories have generally failed to account for the phenomena they were invented to explain (Levin & Rappaport-Hovav, 2005).

The complexity of the empirical situation has made study interpretation challenging. Few studies carefully distinguish between Levin classes, discourse structures, or generalization across them. Many still assume a universal “subject bias” or “first-mention bias” in pronoun comprehension, despite mixed evidence outside topic-preserving discourse structures (e.g., Arnold, 2023; Blything et al., 2021, 2022). Ironically, much of this evidence comes from production studies using sentence-continuation tasks (e.g., Frederiksen & Mayberry, 2022; Hoek et al., 2021; Kehler & Rohde, 2019; Zhan et al.,

2020). Participants complete sentences either before (“Al frightened Bart because...”) or after (“Al frightened Bart because he...”) the pronoun. Production biases (who they reference) and comprehension biases (who they think the pronoun refers to) are inferred from coders’ interpretations rather than direct participant judgments. These studies at least partly measure the judgements of the researchers (inter-coder agreement rarely passes 90%), leaving some ambiguity about the results (see Ariel, 2013 and Arnold, 2023 for a discussion of concerns about this method; also see Weatherford & Arnold, 2021, for a different method designed to address biases in character reference).

To establish a general subject bias in comprehension, studies must systematically test a representative verb sample across Levin classes, discourse structures, and comprehension rather than production. The only study approaching this, Hartshorne & Snedeker (2013), examined pronoun comprehension in explanation discourses across ~1,000 high-frequency verbs from different Levin classes, finding an overall object bias.

Despite progress in identifying pronoun resolution factors, the empirical picture remains unclear. Worse, while researchers increasingly acknowledge that findings in one language may not generalize, studies remain largely English-centric (cf. De la Fuente et al., 2016; Hartshorne et al., 2013; Rudolph & Forsterling, 1997). To begin addressing this gap, we conduct a tightly controlled study in Turkish, investigating interactions between word order, null/overt pronouns, and Levin verb classes in explanation discourses, focusing solely on comprehension. In the remainder of this Introduction, we motivate the specific manipulations we consider, particularly with respect to our language of study: Turkish.

Null vs Overt Pronouns

Turkish allows the dropping of both subject and object arguments, with the missing argument being dubbed a “null pronoun.” Many authors have posited that in languages that have null pronouns, speakers use them preferentially highly-salient entities (such as the topic of the previous clause), and that listeners likewise preferentially resolve null pronouns to the previous topic (etc.) (Ariel, 1990; Gundel et al., 1993). Thus, using an overt pronoun would indicate a topic shift — something that could be hard-wired into the grammar, or which could be derived through Gricean principles (Ariel, 1990; Gundel et al., 1993). Such arguments have been made for Turkish (Enç, 1989; Turan, 1997). Kerslake (1987), however, argues that although the subject tends to be re-mentioned as a null pronoun, it may appear as an overt pronoun when it has the contrastive stress.

(e.g., Rudolph & Forsterling, 1997). This assumption was not systematically tested until Hartshorne (2014), who directly compared the two phenomena in eight large experiments, finding essentially no relationship between the two. Critically, many contextual factors that systematically affect social attribution and were thus widely assumed to also affect pronoun comprehension turned out not to.

¹ Note also that a great deal of confusion has been caused by that there are at least two, entirely distinct, phenomena that are both known as “implicit causality”: One is the pronoun bias effect we discuss here, while the other involves inferences about social interactions and does not involve pronouns at all. For some decades, it was assumed that these two phenomena were identical, and literature reviews frequently generalized from one to the other

Regardless, it is unclear whether null pronouns preferentially refer to the previous topic across all discourse structures and verbs, whether in Turkish or any other language. Indeed, Turan (1997) argues that null pronouns generally must refer to the previous subject except in a small number of contexts, such as when the previous verb is an EXPERIENCER, in which case a null pronoun preferentially refers to the previous object and an overt pronoun refers to the previous subject. Turan does not consider a role for discourse structure in the sense of Kehler (2002).² Erguvanlı-Taylan (1986) suggests that the pattern may change depending on the clause type: null pronouns can be used to refer to the previous object if the two clauses are conjoined, but null pronouns can only refer to the previous subject if the second clause is subordinate. Since systematically comparing all discourse structures is beyond the scope of this initial study, we focus on contrasting null and overt pronouns in explanation (implicit causality) discourses.

Information Structure and Word Order

Researchers have glossed the phenomena discussed in the previous section in different ways, variably arguing that pronouns (or, in languages that have them, specifically null pronouns) preferentially refer to the subject, topic, or first-mentioned entity in the previous clause (Ariel, 1990; Arnold, 1998; Corbett & Chang, 1983; Cowles, Walenski, & Kluender, 2007; Crawley et al., 1990; Grosz et al., 1983; Kaiser, 2011; Smyth, 1994). As already noted, these are difficult to distinguish in English, though there is some evidence of separable effects of all three, especially from languages that more readily distinguish them, such as Japanese or Finnish (Crawley & Stevenson, 1990; Cowles et al., 2007; Gordon & Chan, 1995; Järvikivi et al. 2005; Foraker & McElree, 2007; Kaiser & Trueswell 2008; Kaiser, 2011; Walker et al., 1994). Again, though, these studies do not allow us to determine whether the results are specific to particular discourse structures, much less across the handful of languages that have been studied.

Turkish provides another opportunity to contrast subjecthood with topichood and order-of-mention. Turkish is an SOV language that allows variation of the constituents in six possible word order configurations. This means, unlike English, it is straightforward in Turkish to dissociate first-mention for subject. Note that word order is not, however, separable from topichood: the topic occupies sentence initial position while the preverbal position is the focus position (Erguvanlı-Taylan 1986; Erkü 1983; İşsever 2003 among others).

Özge and Evcen (2020) investigated the role of word order on pronoun comprehension in Turkish, contrasting SOV (1) with OSV (2) word orders. In Turkish, the topic is the first argument, while the preverbal argument is the focus.

- | | | | |
|-----|--|------------------------|------------------|
| (1) | <i>Bahar</i> | <i>Ceren'i ara-di;</i> | |
| | Bahar-NOM | Ceren-ACC | call-PST.3SG |
| | <i>sadece</i> | <i>o</i> | <i>müsait-ti</i> |
| | only | she | free-COP.PST.3SG |
| | 'Bahar called Ceren because only she was free.' | | |
| (2) | <i>Ceren'i Bahar</i> | <i>ara-di;</i> | |
| | Ceren-ACC | Bahar-NOM | call-PST.3SG |
| | <i>sadece</i> | <i>o</i> | <i>müsait-ti</i> |
| | only | she | free-COP.PST.3SG |
| | 'As for Ceren, Bahar called her; only she was free.' | | |

They found a slight but significant effect, with participants more frequently resolving the pronoun to the previous subject in the OSV order, suggesting a slight preference to resolve pronouns to the focus than the topic. This result was replicated with null pronouns, and no significant difference was observed between null and overt pronouns. This contrasts with previous reports that null pronouns preferentially refer to the topic (Enç, 1989; Kameyama 1985; Turan 1995; Prince 1999) or are unaffected by word order (Turing, 1997). However, Özge and Evcen considered only consequence discourses, and did not systematically consider verb class.

Implicit Causality and Levin Verb Classes

In a seminal *tour de force* analysis, Levin (1993) showed that English verbs form classes based on their argument structure behavior (whether they allow direct objects, participate in the dative alternation, etc.; for details, see Levin, 1993). Critically, it appears that verbs in the same “class” are systematically related in terms of semantics (Ambridge et al., 2013; Hartshorne et al., 2014; Jackendoff, 1990; Levin, 1993; Pinker, 1989).

In a series of papers, Hartshorne and colleagues showed that pronoun interpretation in explanation and consequence discourses (also known as “implicit causality” and “implicit consequentiality” sentences) is reliably modulated by the Levin verb class of the first clause (Hartshorne & Snedeker, 2013; Hartshorne et al., 2015). Importantly, Levin verb class was far more predictive than thematic roles, which in most cases did no better than chance. For instance, it is widely claimed that “psych” verbs where the subject bears the STIMULUS role (*frighten, appeal to, compel*) are subject-biased in explanation discourses (implicit causality) and object-biased in consequence discourses (implicit consequentiality), with the reverse being true of psych verbs where the object bears the STIMULUS role (*fear, blame, care about, believe*) (Crinean & Garnham, 2006; Pickering & Majid, 2007; Rudolf & Forsterling, 1997; Solstad & Bott, 2022). However, this turns out to be an artifact of the fact that most of the verbs tested have been transitive emotion verbs (*frighten, fear*; Levin classes 31.1 and 31.2); psych verbs belonging to other Levin classes behave differently

² Note that here and elsewhere, we reference thematic roles like EXPERIENCER as a useful shorthand, without taking a stance on their cognitive reality (see Levin & Rappaport Hovav, 2005).

(Hartshorne & Snedeker, 2013; Hartshorne et al., 2015). Hartshorne and colleagues (2013) further showed that the two classes of emotion verbs behave similarly in Japanese, Mandarin, Russian, Spanish, Finnish, Dutch, and Italian, at least in explanation (implicit causality) discourses. There has been no work to date on this topic in Turkish.

Experiment

As reviewed above, it is generally believed that in languages with null pronouns, null pronouns preferentially refer to the previous subject while overt pronouns do not. However, prior work has not carefully controlled for Levin verb class or discourse structure (in the sense of Kehler, 2002), factors which are known to modulate or even reverse pronoun preferences. There is some evidence that subject-preferences for pronoun resolution are stronger in SO sentences relative to OS in those languages that allow both. However, because prior work does not systematically control Levin verb class or discourse structure, it is quite possible that those findings are an artifact of the stimuli used. Because Turkish has flexible word order and uses both null and overt pronouns, it provides a convenient opportunity to begin disentangling these issues.

We conducted an initial study of whether and how referential form, word order, and implicit causality interact in the interpretation of ambiguous anaphors in Turkish, using a forced choice comprehension study modeled after Hartshorne and Snedeker (2013). However, Turkish provides a useful opportunity to consider word order, because both SOV and OSV are reasonably natural. It also, unlike English, allows us to contrast null pronouns with overt pronouns.

Investigating the role of Levin verb class is made complex by the fact that English and Czech (Pala & Horák, 2008) are the only languages to have been comprehensively analyzed according to verb argument structure patterns (for initial forays in other languages, see: Aparicio et al., 2008; Estarrona et al., 2016; Liu, 2020; Pradet et al., 2014). Thus, following Hartshorne et al. (2013), we focus on emotion verbs, which are relatively easy to identify and have shown reasonably uniform behavior in languages spanning four language families (Hartshorne et al., 2013). Note, though, that strictly speaking we cannot at this point be certain they are all in the same Levin verb class, which would require a comprehensive study of argument structure alternations in Turkish. Note also that emotion verbs is a somewhat more narrow category than verbs of psychological state (“psych verbs”): the latter includes several Levin verb classes, such as verbs of cognition (*think, believe*), which are distinct from emotion verbs both in terms of syntax and also (at least in English) affect pronoun interpretation differently (Hartshorne & Snedeker, 2013). In order to make results more interpretable, we focus on a single discourse structure (explanation discourses). We leave comparison of different discourse structures to future work. To our knowledge, this is the first study in any language to test the interaction of implicit causality with word order.

Because this project brings together several different streams of research, existing theories do not necessarily address all the manipulations tested here. The exception is Turan (1997), who explicitly predicts that a) null pronouns will be interpreted as referring to the EXPERIENCER of the previous clause whereas overt pronouns will refer to the STIMULUS, and b) there will be no effect of word order. In contrast, researchers working in the implicit causality literature would likely predict that both null and overt pronouns will refer to the STIMULUS, though this may be attenuated somewhat by an overall subject bias for null pronouns (Fedele & Kaiser, 2015; Hartshorne et al., 2013; Ueno & Kehler, 2016; but see Román, 2020). These accounts make no clear predictions about our word-order manipulation.

Methods

All materials, data, and analysis code for this and all the following experiments are available at <https://osf.io/jn9fh/>.

Participants: 136 native speakers of Turkish who were all undergraduate students in Turkey completed the study. No participants were excluded because none met our exclusion criterion of choosing the same referent across all conditions, which would be a sign of not paying attention. 29 participants were randomly assigned to the first set, 37 participants to the second set, 32 participants to the third set, and 38 participants to the final set. The imbalances in sample sizes across the four lists were due to some participants starting the experiment but not completing it.

Materials & Procedures: We designed a 2x2x2 study where we manipulated the word order (SOV, OSV), referential form (null vs overt pronoun) between-participants and verb type (experiencer-subject, experiencer-object verbs) within-participants. We selected the emotion verbs that exhibited the strongest implicit causality effects in Hartshorne and Snedeker (2013) and translated them into Turkish. We eliminated causatives and/or other case markings (genitive, ablative), so we restricted our verb list to accusative marked objects and animate subject-animate object possible verbs only. From these, we selected 8 fear-type verbs (onayla-/favor, arzula-/desire, beğen-/like, sev-/love, affet-/pardon, kıskan-/envy, aşağıla-/d disdain, yargıla-/judge) and 8 frighten-type verbs (büyüle-/fascinate, cezbet-/attract, etkile-/dazzle, sık-/bore, kır-/offend, üz-/upset, yarala-/hurt, örsele-/mistreat). A sample item for each condition was presented in (3) and (4):

(3) SOV, Null/Overt pronoun, Exp-subject-/object verb

Bahar	Ceren’i	arzulu-yor/büyülü-yor
Bahar-NOM	Ceren-ACC	desire/dazzle-PROG.3SG
çünkü	(o)	dakmuk
because	(she)	dakmuk

‘Bahar desires/dazzles Ceren because she is dakmuk.’

(4) **OSV, Null/Overt pronoun, Exp-subject/-object verb**

Bahar’ı	Ceren	arzulu-yor/büyülü-yor
Bahar- ACC	Ceren-NOM	desire/dazzle-PROG-3SG
çünkü	(o)	dakmuk
because	(she)	dakmuk

‘Ceren desires/dazzles Bahar because she is dakmuk.’

We presented participants with a sentence composed of a main clause and a subordinate clause conjoined with ‘because’. The main clause was composed of two referents with [+human] and [+female] features for the subject and the object entity (e.g., Sally frightens Mary). The subordinate clause was constructed with an ambiguous pronoun and a non-word adjectival predicate ‘dax’ (e.g., ...because she is dax). We asked participants to decide who the referent for the non-word is. This made participants resolve the ambiguous pronoun towards the subject or the object without explicitly asking them who the pronoun refers to.

We created four counterbalanced and pseudo-randomized lists with 16 critical (8 experiencer-subject, 8 experiencer-object) and 24 filler items. The filler items included other verb types such as physical touch action verbs, reciprocal verbs and benefactives. We also counterbalanced the order of the answer options (*Bahar, Ceren*) for the question (*Who is dakmuk?*) so that each referent appeared equally often.

We collected offline responses from participants through a web-based survey tool. Each trial and the following question were presented one at a time to prevent participants from making comparisons between their judgments and changing their answers. Participants used their phones as small groups in a classroom setting, and they all completed the task individually. The experiment took approximately 15 minutes to complete.

Results

First, to check the validity of our groups, we conducted a pairwise t-test to compare subject preference scores between the two verb groups. The results indicated a significant difference between experiencer-object verbs and experiencer-subject verbs ($t(260) = -8.23, p < .001$). Additionally, we conducted one-sample t-tests to compare each verb group to chance. The results revealed that experiencer-object verbs exhibited a strong subject bias, with scores significantly greater than chance ($M = 72.88, SD = 0.22, t(135) = 12.01, p < .001$). In contrast, experiencer-subject verbs did not show a significant bias toward either subject or object ($M = 48.16, SD = 0.27, t(135) = -0.79, p = .40$).

Second, we submit the data to a partially-Bayesian generalized mixed effects logistic regression with Wishart

priors on the covariance matrix for random effects using the *blme* package with bobyqa optimization in order to improve convergence and avoid issues with singularity (Chung et al., 2013). Models were constructed to predict subject preference from contrast-coded fixed effects of Word Order (SOV, OSV), Referential Form (null, overt), and Verb Type (experiencer-subject, experiencer-object), including their interactions. We began with a model that included maximal effects structure and refined the model removing non-contributory variance components until no further improvements were observed via likelihood ratio tests (Bates et al., 2015). Our final model included as main effects all three fixed effects and the interaction of Verb Type and Referential Form and Word Order and Referential Form (see [anonymized link] for model selection details). Random effects include random intercepts of subject and item as well as, and random slopes of Verb Type by subject and Word Order by item. We report the output of the best-fitting model regression coefficient, standard error, Wald’s z-value, and p-value.³

The model revealed a significant main effect of Verb Type ($\beta = 1.21, SE = 0.18, z = 6.62, p < .001$) such that subject responses were produced significantly more often for experiencer-object verbs ($M = 72.88; 95\% CI[69.11, 76.65]$) than experiencer-subject verbs ($M = 48.16; 95\% CI[43.57, 52.74]$) and an effect of Word Order ($\beta = 0.52, SE = 0.24, z = 2.49, p = .02$), reflecting a larger subject preference in OSV order ($M = 65.08; 95\% CI[60.51, 69.65]$) than in SOV order ($M = 55.68; 95\% CI[50.61, 60.75]$). There was also a significant interaction between Verb Type and Referential Form ($\beta = 0.61, SE = 0.28, z = 2.16, p = .03$). The difference between the two levels of referential form is not statistically

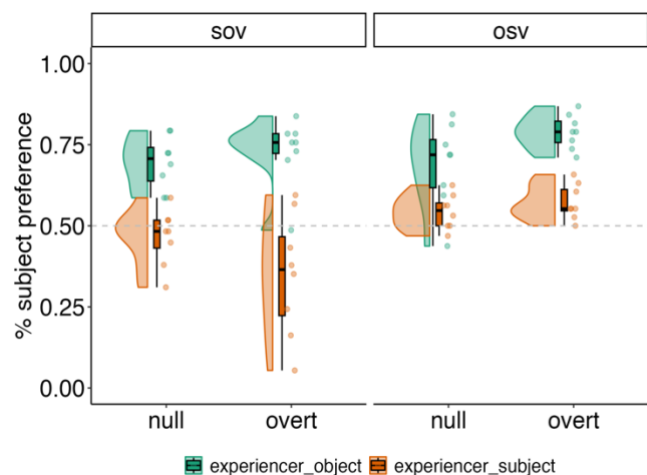


Figure 1: Proportion of choice of subject for each Word Order (SOV, OSV), Referential Form (null, overt) and Verb Type (experiencer-subject, experiencer-object). Each dot represents an item.

³ A model with all three two-way interactions fit slightly better ($p = .04$). However, neither of the other two two-way interactions were individually significant in either a log-likelihood test or using Wald’s z. Thus, for simplicity of description, we omit them. For

details on this model – which supports the same conclusions as the one described here – see <https://osf.io/jn9fh/> for model selection criteria.

significant at either level of verb type, although the direction and magnitude of the effect are different: For the experiencer-subject verbs, subject preference did not differ between the null pronoun condition ($M=50.61$; 95% CI[44.06, 57.16]) and the overt pronoun condition ($M=46.16$; 95% CI[39.68, 52.64]) ($p = .3$) whereas for the experiencer-object verbs, the overt pronoun condition ($M=76.01$; 95% CI[71.42, 80.57]) resulted in more subject preference than the null pronoun condition ($M=69.05$; 95% CI[62.8, 76.31]), though the effect was not significant ($\beta = -0.4$, $SE = 0.23$, Wald's $z = -1.76$, $p=.07$).⁴ There were no other significant effects.

General Discussion

Our aim in this study was to investigate how the factors like information structure encoded in word order variation (SOV versus OSV), the type of referential form (null versus overt pronouns), and verb semantics (implicit causality) influence the interpretation of pronouns in Turkish. In our study, participants read an utterance constructed with an emotion verb, followed by another clause connected with a causal connective 'because'. The second clause ended with a nonsense adjectival predicate 'dakmuk'. Given a binary choice between the two referents (Subject, Object) of each utterance, the participants determined the antecedent of this adjectival predicate.

Across all conditions, subjects were more likely to resolve the pronoun to the subject for experiencer-object verbs than experiencer-subject verbs. There was a slightly stronger tendency to resolve pronouns to the previous subject in OSV order relative to SOV, which numerically was more pronounced for experiencer-subject verbs. The only interaction was between verb-type and referential (pronoun) form, though it was small and not entirely clear. Numerically, at least, it appears that experiencer-object verbs showed a stronger subject preference with overt pronouns than null pronouns, though this was not significant.

These results contrast clearly with the predictions of Turan (1997), who suggests the opposite effect for verb-type; an interaction of verb-type with reference type, which we did not find; and no effect of word order, which we did find. In retrospect, it seems likely that Turan's (1997) predictions about verb-type were influenced by the nature of her data. Turan's (1997) study used examples from naturally-occurring written data, where the animacy across the verb type was not controlled. In her examples, the experiencer was animate while the stimulus was inanimate, which may have biased participants to expect pronouns to refer to the (animate) EXPERIENCERS. In contrast, our study controlled animacy. We do not currently have a hypothesis as to why Turan's (1997) investigation suggested that reference type interacts

with verb-type. We return to the word-order results below.

In contrast, our data are largely consistent with what would be predicted from the implicit causality literature: both overt and null pronouns are resolved to the previous STIMULUS (this literature made no clear prediction about word order, which we return to below). The primary caveat is that there was an overall subject bias (in most conditions, the experiencer-subject verbs are roughly equi-biased rather than being strongly object-biased), which is not quite what implicit causality research would tend to suggest. However, across implicit causality studies, there is quite a bit of variability in whether studies show overall subject preferences, overall object preferences, or no preference [for instance, compare Hartshorne & Snedeker (2012) with Hartshorne et al. (2015)]. The reasons remain unclear and require further investigation.

The existence of a word-order effect was not predicted by the implicit causality literature (which makes no clear predictions) and contrasts with Turan's (1997) explicit predictions of no effect. The actual effect we found (a slight increase in subject preference in OSV) was thus unexpected. As reviewed above, in Turkish, the first argument is typically analyzed as the topic while the second is analyzed as the focus. Thus, our results are consistent with a slight preference for resolving pronouns to the previous focus rather than topic. We are unaware of any analyses that would predict this and quite a few that would predict the opposite (e.g., Ariel, 2013; Grosz, Joshi, & Weinstein, 1983; Hoffman, 1998). One possible explanation is an over-additive interaction of subjecthood and focus-hood. However, the effect is relatively small, so for the moment we would not want to make too much of it. However, it does underscore the complexity of pronoun resolution and the degree to which classic theories lack mechanisms to account for that complexity.

Anaphora have been a core topic of study for the language sciences for decades. It has become increasingly clear that simple theories are inadequate to capture the complexity of the phenomenon. Unfortunately, the discovery of this complexity also makes it difficult to interpret many earlier studies, which do not control for factors that are now known to be critical. Here, we take advantage of the affordances of Turkish to investigate the effects of reference type (null vs. overt pronouns) and word order, focusing on two verb-types and one discourse structure. As one might expect at this point, we find some unexpected results. Clarifying the factors underlying pronoun resolution will require many such studies.

⁴ We conducted separate exploratory analyses, including the fixed effect of Emotional Valence. Practical considerations, such as the limited number of items, restricted more rigorous testing. These findings are very preliminary and require further investigation (see [anonymized link] for details).

References

- Ambridge, B., Pine, J. M., Rowland, C. F., Chang, F., & Bidgood, A. (2013). The retreat from overgeneralization in child language acquisition: Word learning, morphology, and verb argument structure. *Wiley Interdisciplinary Reviews: Cognitive Science*, 4(1), 47–62.
- Aparicio, J., Taulé, M., & Martí, M. A. (2008). AnCora-Verb: A lexical resource for the semantic annotation of corpora. In N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, & D. Tapias (Eds.), *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*. European Language Resources Association (ELRA).
- Ariel, M. (1990). *Accessing noun-phrase antecedents (RLE Linguistics B: Grammar)* (1st ed.). Routledge.
- Ariel, M. (2013). Centering, accessibility and the next mention. *Theoretical Linguistics*, 39(1-2), 39-58.
- Arnold, J. E. (1998). *Reference form and discourse patterns*. [Doctoral dissertation, Stanford University]
- Arnold, J. E. (2023). *Pronoun comprehension*. In *The Routledge Handbook of Pronouns* (pp. 120–134). Routledge.
- Au, T. K. (1986). A verb is worth a thousand words: The causes and consequences of interpersonal events implicit in language. *Journal of Memory and Language*, 25(1), 104-122.
- Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., ... & Bolker, M. B. (2015). Package 'lme4'. *Convergence*, 12(1), 2.
- Blything, L. P., Iraola Azpiroz, M., Allen, S., Hert, R., & Järvikivi, J. (2022). The influence of prominence cues in 7- to 10-year-olds' pronoun resolution: Disentangling order of mention, grammatical role, and semantic role. *Journal of Child Language*, 49(5), 930–958.
- Blything, L. P., Järvikivi, J., Toth, A. G., & Arnhold, A. (2021). The influence of focus marking on pronoun resolution in dialogue context. *Frontiers in Psychology*, 12, 684639.
- Brown, R., & Fish, D. (1983). The psychological causality implicit in language. *Cognition*, 14(3), 237-273.
- Chung, Y., Rabe-Hesketh, S., Dorie, V., Gelman, A., & Liu, J. (2013). A nondegenerate penalized likelihood estimator for variance parameters in multilevel models. *Psychometrika*, 78(4), 685–709.
- Corbett, A. T., & Chang, F. R. (1983). Pronoun disambiguation: Accessing potential antecedents. *Memory & Cognition*, 11(3), 283–294.
- Cowles, H. W., Walenski, M., & Kluender, R. (2007). Linguistic and cognitive prominence in anaphor resolution: Topic, contrastive focus, and pronouns. *Topoi*, 26, 3-18.
- Crawley, R. A., & Stevenson, R. J. (1990). Reference in single sentences and in texts. *Journal of Psycholinguistic Research*, 19, 191-210.
- Crawley, R. A., Stevenson, R. J., & Kleinman, D. (1990). The use of heuristic strategies in the interpretation of pronouns. *Journal of Psycholinguistic Research*, 19, 245-264.
- Crinean, M., & Garnham, A. (2006). Implicit causality, implicit consequentiality, and semantic roles. *Language and Cognitive Processes*, 21(5), 636-648.
- De la Fuente, I., Hemforth, B., Colonna, S., & Schimke, S. (2016). The role of syntax, semantics, and pragmatics in pronoun resolution: A cross-linguistic overview. In P. Bouillon, P. Sturgeon, & H. Isahara (Eds.), *Empirical perspectives on anaphora resolution* (pp. 11-31). Springer.
- Enç, M. (1989). Pronouns, licensing, and binding. *Natural Language & Linguistic Theory*, 7, 51–92.
- Erguvanli-Taylan, E. (1986). Pronominal vs. zero representation of anaphora in Turkish. In D. I. Slobin & K. Zimmer (Eds.), *Studies in Turkish Linguistics* (pp. 206–233). John Benjamins.
- Erk, F. (1983). *Discourse pragmatics and word order in Turkish* [Doctoral dissertation, University of Minnesota].
- Estarrona, A., Aldezabal, I., Daz de Ilarraza, A., & Aranzabe, M. J. (2016). A methodology for the semiautomatic annotation of EPEC-RolSem, a Basque corpus labeled at predicate level following the PropBank-VerbNet model. *Digital Scholarship in the Humanities*, 31(3), 470-492.
- Fedele, E., & Kaiser, E. (2015). Resolving null and overt pronouns in Italian: An experimental investigation of syntax–semantics interactions. In *TLS: Proceedings of the 15th Texas Linguistic Society* (pp. 53-72). University of Texas at Austin.
- Forstl, E. C., Garnham, A., & Manouilidou, C. (2011). Implicit causality bias in English: A corpus of 300 verbs. *Behavior Research Methods*, 43, 124-135.
- Foraker, S., & McElree, B. (2007). The role of prominence in pronoun resolution: Active versus passive representations. *Journal of Memory and Language*, 56(3), 357-383.
- Frederiksen, A. T., & Mayberry, R. I. (2022). Pronoun production and comprehension in American Sign Language: The interaction of space, grammar, and semantics. *Language, Cognition and Neuroscience*, 37(1), 80-102.
- Garnham, A., Vorthmann, S., & Kaplanova, K. (2021). Implicit consequentiality bias in English: A corpus of 300+ verbs. *Behavior Research Methods*, 53, 1530-1550.
- Garvey, C., & Caramazza, A. (1974). Implicit causality in verbs. *Linguistic Inquiry*, 5(3), 459-464.
- Givn, T. (1976). Topic, pronoun, and grammatical agreement. In C. N. Li (Ed.), *Subject and topic* (pp. 149–188). Academic Press.
- Gordon, P. C., & Chan, D. (1995). Pronouns, passives, and discourse coherence. *Journal of Memory and Language*, 34(2), 216-231.
- Grosz, B. J., Joshi, A. K., & Weinstein, S. (1983). Providing a unified account of definite noun phrases in discourse. In *Proceedings of the 21st Annual Meeting of the Association for Computational Linguistics* (pp. 44–50). Association for Computational Linguistics.
<https://doi.org/10.3115/981311.981320>

- Gundel, J. K., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language*, 69(2), 274-307.
- Hartshorne, J. K. (2014). What is implicit causality?. *Language, Cognition and Neuroscience*, 29(7), 804-824.
- Hartshorne, J. K., & Snedeker, J. (2013). Verb argument structure predicts implicit causality: The advantages of finer-grained semantics. *Language and Cognitive Processes*, 28(10), 1474-1508.
- Hartshorne, J. K., Bonial, C., & Palmer, M. (2014). The VerbCorner Project: Findings from Phase 1 of crowdsourcing a semantic decomposition of verbs. In K. Toutanova & H. Wu (Eds.), *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)* (pp. 397-402). Association for Computational Linguistics.
- Hartshorne, J. K., O'Donnell, T. J., & Tenenbaum, J. B. (2015). The causes and consequences explicit in verbs. *Language, Cognition and Neuroscience*, 30(6), 716-734.
- Hartshorne, J. K., Sudo, Y., & Uruwashi, M. (2013). Are implicit causality pronoun resolution biases consistent across languages and cultures?. *Experimental Psychology*, 60(3), 179-196.
- Hoek, J., Kehler, A., & Rohde, H. (2021). Pronominalization and expectations for re-mention: Modeling coreference in contexts with three referents. *Frontiers in Communication*, 6, 674126.
- Hoffman, B. (1998). Word order, information structure, and centering in Turkish. In M. A. Walker, A. K. Joshi, & E. F. Prince (Eds.), *Centering theory in discourse* (pp. 251-272). Oxford University Press.
- İşsever, S. (2003). Information structure in Turkish: The word order-prosody interface. *Lingua*, 113(11), 1025-1053.
- Jackendoff, R. S. (1992). *Semantic structures*. MIT Press.
- Järvikivi, J., Van Gompel, R. P., Hyönä, J., & Bertram, R. (2005). Ambiguous pronoun resolution: Contrasting the first-mention and subject-preference accounts. *Psychological Science*, 16(4), 260-264.
- Kaiser, E. (2011). Focusing on pronouns: Consequences of subjecthood, pronominalization, and contrastive focus. *Language and Cognitive Processes*, 26(10), 1625-1666.
- Kaiser, E., & Trueswell, J. C. (2008). Interpreting pronouns and demonstratives in Finnish: Evidence for a form-specific approach to reference resolution. *Language and Cognitive Processes*, 23(5), 709-748.
- Kameyama, M. (1985). *Zero anaphora: The case of Japanese* [Doctoral dissertation, Stanford University].
- Kehler, A. (2002). *Coherence, reference, and the theory of grammar*. CSLI Publications.
- Kehler, A., & Rohde, H. (2019). Prominence and coherence in a Bayesian theory of pronoun interpretation. *Journal of Pragmatics*, 154, 63-78.
- Kehler, A., Kertz, L., Rohde, H., & Elman, J. L. (2008). Coherence and coreference revisited. *Journal of Semantics*, 25(1), 1-44.
- Kerslake, C. (1987). Noun phrase deletion and pronominalization in Turkish. In H. E. Boeschoten & L. T. Verhoeven (Eds.), *Studies on Modern Turkish* (pp. 91-104). Tilburg University Press.
- Levin, B. (1993). *English verb classes and alternations: A preliminary investigation*. University of Chicago Press.
- Levin, B., & Rappaport Hovav, M. (2005). *Argument realization*. Cambridge University Press.
- Liu, M. (2020). *The Construction and Annotation of a Semantically Enriched Database: The Mandarin VerbNet and Its NLP Applications*. In Q. Su, & W. Zhan (Eds.), *From Minimal Contrast to Meaning Construct: Corpus-based, Near Synonym Driven Approaches to Chinese Lexical Semantics* (pp. 257-272). Springer Nature Singapore Pte Ltd.
- Özge, D. & Evcen, E. (2020). Referential form, word order and emotional valence in Turkish pronoun resolution in physical contact events. In D. Zeyrek & U. Özge (Ed.), *Discourse Meaning: The View from Turkish* (pp. 165-186). Berlin, Boston: De Gruyter Mouton.
- Pala, K., & Horák, A. (2008). Can complex valency frames be universal? In P. Sojka & A. Horák (Eds.), *Proceedings of Recent Advances in Slavonic Natural Language Processing, RASLAN 2008* (pp. 41-48).
- Pickering, M. J., & Majid, A. (2007). What are implicit causality and consequentiality? *Language and Cognitive Processes*, 22(5), 780-788.
- Pinker, S. (1989). *Learnability and cognition: The acquisition of argument structure*. MIT Press.
- Pradet, Q., Danlos, L., & De Chalendar, G. (2014). Adapting VerbNet to French using existing resources. In *LREC'14-Ninth International Conference on Language Resources and Evaluation*. European Language Resources Association (ELRA).
- Prince, E. F. (1998). Subject-Prodrop in Yiddish. In P. Bosch & R. A. van der Sandt (Eds.), *Focus: linguistic, cognitive, and computational perspectives* (pp. 82-104). Cambridge: Cambridge University Press.
- Román, D. C. (2020). *Implicit causality, pronominal form and anaphora resolution in Spanish* [Doctoral dissertation, University of Alberta].
- Rudolph, U., & Försterling, F. (1997). The psychological causality implicit in verbs: A review. *Psychological Bulletin*, 121(2), 192-218.
- Smyth, R. (1994). Grammatical determinants of ambiguous pronoun resolution. *Journal of Psycholinguistic Research*, 23(3), 197-229.
- Solstad, T., & Bott, O. (2022). On the nature of implicit causality and consequentiality: The case of psychological verbs. *Language, Cognition and Neuroscience*, 37(10), 1311-1340.
- Song, J., & Kaiser, E. (2024). Effects of referential structure on pronoun interpretation. *Language, Cognition and Neuroscience*, 39(1), 98-117.
- Stevenson, R. J., Crawley, R. A., & Kleinman, D. (1994). Thematic roles, focus, and the representation of events. *Language and Cognitive Processes*, 9(4), 519-548.

- Turan, Ü. D. (1995). Subject and object in Turkish discourse: A centering analysis [Doctoral dissertation, University of Pennsylvania].
- Turan, Ü. D. (1997). Ranking forward-looking centers in Turkish: Universal and language-specific properties. In M. Walker, E. Prince, & A. Joshi (Eds.), *Centering theory in discourse* (pp. 139–160). Clarendon Press.
- Ueno, M., & Kehler, A. (2016). Grammatical and pragmatic factors in the interpretation of Japanese null and overt pronouns. *Linguistics*, *54*(6), 1165-1221.
- Walker, M., Iida, M., & Cote, S. (1994). Japanese discourse and the process of centering. *Computational Linguistics*, *20*(2), 193–231.
- Weatherford, K. C., & Arnold, J. E. (2021). Semantic predictability of implicit causality can affect referential form choice. *Cognition*, *214*, 104759.
- Zhan, M., Levy, R., & Kehler, A. (2020). Pronoun interpretation in Mandarin Chinese follows principles of Bayesian inference. *PLoS One*, *15*(8), e0237012.