University of Groningen

Reasoning about alternative forms is costly

Vogelzang, Margreet; Foppolo, Francesca; Guasti, Maria Teresa; van Rijn, Hedderik; Hendriks, Petra

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Version created as part of publication process; publisher's layout; not normally made publicly available

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):
Vogelzang, M., Foppolo, F., Guasti, M. T., van Rijn, H., & Hendriks, P. (2020). Reasoning about alternative forms is costly: The processing of null and overt pronouns in Italian using pupillary responses. *Discourse Processes*, *57*(2), 158-183. https://doi.org/10.1080/0163853X.2019.1591127

Reasoning about alternative forms is costly: The processing of null and overt pronouns in Italian using pupillary responses

Margreet Vogelzang (corresponding author)

Center for Language and Cognition Groningen, University of Groningen

P.O. Box 716, 9700 AS Groningen, The Netherlands

Now at Cluster of Excellence "Hearing4all" &

Institute of Dutch Studies, University of Oldenburg

Ammerländer Heerstraße 114-118, 26111 Oldenburg, Germany

Email: margreet.vogelzang@uni-oldenburg.de

Francesca Foppolo

Department of Psychology, University of Milano-Bicocca

Piazza dell'Ateneo Nuovo 1, 20126 Milan, Italy

Email: francesca.foppolo@unimib.it

Maria Teresa Guasti

Department of Psychology, University of Milano-Bicocca

Piazza dell'Ateneo Nuovo 1, 20126 Milan, Italy

Email: mariateresa.guasti@unimib.it

Hedderik van Rijn

Department of Psychology, University of Groningen

Grote Kruisstraat 2/1, 9712 TS Groningen, The Netherlands

Email: d.h.van.rijn@rug.nl

Petra Hendriks

Center for Language and Cognition Groningen, University of Groningen

P.O. Box 716, 9700 AS Groningen, The Netherlands

Email: p.hendriks@rug.nl

Reasoning about alternative forms is costly: The processing of null and overt pronouns in Italian using pupillary responses

Abstract

Different words generally have different meanings. However, some words seemingly share similar meanings. An example are null and overt pronouns in Italian, which both refer to an individual in the discourse. Is the interpretation and processing of a form affected by the existence of another form with a similar meaning? With a pupillary response study, we show that null and overt pronouns are processed differently. Specifically, null pronouns are found to be less costly to process than overt pronouns. We argue that this difference is caused by an additional reasoning step which is needed to process marked overt pronouns, but not unmarked null pronouns. A comparison to data from Dutch, a language with overt but no null pronouns, demonstrates that Italian pronouns are processed differently from Dutch pronouns. These findings suggest that the processing of a marked form is influenced by alternative forms within the same language, making its processing costly.

Keywords: Ambiguity resolution; null subjects; pronouns; pupil dilation; cognitive effort.

1 Introduction: Reasoning about alternative forms

A fundamental property of language is reference. Words and phrases in a language typically refer to things, people, thoughts, situations, and events in the world. Processing such referring expressions and understanding what they refer to is essential for successful communication. Commonly used referring expressions in English are full noun phrases (NPs) such as *the boy* and personal pronouns such as *he*. Here we will discuss how the understanding of referring expressions is influenced by alternative forms available in the language.

1.1 Marked versus unmarked forms

In general, if there are two forms available in a language that share the same referential space and can be used to express similar meanings, one of them is a briefer, more lexicalized and possibly more frequent form, and is said to be *unmarked* (Greenberg, 1966; Horn, 1984). Consequently, the other form is the *marked* form. These two forms differ in meaning, with the unmarked form having an unmarked, stereotypical meaning, and the marked form having a less stereotypical meaning for which the unmarked form is usually not used. This is referred to as Horn's division of pragmatic labor (Blutner, 2000; Hendriks, De Hoop, Kramer, De Swart, & Zwarts, 2010). A similar idea is illustrated by Kiparsky's (1982) notion of partial blocking, according to which some forms have a specific meaning, that blocks other, more general, forms from picking up this meaning. For example, when someone uses the marked or general phrase *X caused Y to die*, one generally assumes that X did not kill Y directly, as otherwise the more specific, unmarked form *killed* would have been used (McCawley, 1978).

Both Horn (1984) and Kiparsky (1982) argue that the meaning of the marked form is derived from the meaning of the unmarked form. This is in line with Grice (1975), who proposed that

3

listeners generally assume speakers to be cooperative, and therefore perform a pragmatic reasoning step called an implicature if the speaker uses a form that is not the most informative, brief, or relevant in the specific context of utterance. Thus, if a speaker uses a marked form instead of an alternative unmarked form, the listener may reason that the speaker wanted to express a meaning that is different, often more specific, than the meaning expressed by the unmarked form. Crucially, this pragmatic reasoning step is only relevant when there are potentially better (i.e., more informative, briefer, more relevant) alternative forms for the used form.

In this paper, we will examine whether the availability of alternative referential forms in a given language affects the meaning and processing of a particular referential form, because an additional reasoning step has to be performed. This will be tested in Italian, a null-subject language which not only allows for full NPs and overt pronouns in subject position, but also for null subjects. We will use pupillary responses to investigate whether overt and null pronoun subjects in Italian are processed in a similar manner. We predict that overt and null pronoun subjects are processed differently, because reasoning about alternative forms is required for the processing of overt pronouns, but not for null subjects. Additionally, a cross-linguistic comparison will be carried out to investigate whether similar forms have different meanings in languages which differ in their availability of alternative forms. To this end, the results from the experimental study in Italian will be compared to the results from Vogelzang et al. (2016) on the processing and interpretation of anaphoric subjects in the non-null-subject language Dutch.

1.2 Anaphoric subjects in Italian

In non-null-subject languages like Dutch and English, a speaker can use a full NP or an overt pronoun (such as *he*) as the subject of a sentence (see 1). In contrast, in null-subject languages like Italian and Spanish, speakers have a third option and can choose between using a full NP, an overt pronoun (such as *lui* 'he'), or a null subject (omitting the subject, '∅'), illustrated by 2.

(1)   ■ a. De jongen eet een appel.

       The boy is eating an apple.

     ■ b. Hij eet een appel.

       He is eating an apple.

(2)   ■ a. Il ragazzo mangia una mela.

       The boy is eating an apple.

     ■ b. Lui mangia una mela.

       He is eating an apple.

     ■ c. ∅ mangia una mela.

       ∅ is eating an apple.

The meaning of an anaphoric subject depends on the discourse it appears in. As can be seen in Figure 1 (left side), overt pronoun subjects in non-null-subject languages like Dutch generally refer to the most prominent referent[1] (a.o., Ariel, 1990; Givón, 1983; Gundel, Hedberg, & Zacharski, 1993), which is often called the discourse topic. Conversely, linguistically more constrained forms such as full NP subjects (e.g., *the boy*) refer to a referent that is less prominent (e.g., new to the context, Ariel, 1990; Chafe, 1994).

- Figure 1 here -

As explained, in the null-subject language Italian an additional anaphoric subject form is available compared to English and Dutch. The availability of this additional null subject may result in a different mapping of forms on the prominence scale (Figure 1, right side). Investigating the interpretational preferences of Italian subject pronouns, Carminati (2002) found that Italian preverbal null subjects prefer an antecedent in subject position. The antecedent in subject position is generally the discourse topic, and as a result null subjects indicate a topic continuation. In contrast, the use of an overt pronoun subject for reference to the topical antecedent may be perceived as an inappropriate repetition (Belletti & Guasti, 2015), much like the overuse of full NPs in English (Gordon, Grosz, & Gilliom, 1993). Overt pronoun subjects prefer reference to a non-topical referent, resulting in a topic shift (Carminati, 2002; replicated by Filiaci, 2010). These antecedent preferences in Italian have been confirmed in several experimental studies (Serratrice, 2007; Tsimpli, Sorace, Heycock, & Filiaci, 2004). Consequently, overt pronoun subjects in Italian have a different function than overt pronoun subjects in languages that lack null pronouns.

Even though the antecedent preferences of null and overt pronoun subjects are known, it is unknown how exactly these referring expressions are processed, and how their meaning is determined. We will experimentally address two questions about the processing of Italian referring expressions. Firstly, we will test the hypothesis that overt pronoun subjects are more effortful to process than null subjects. We hypothesize that Italian null subjects, which often refer to the discourse topic, are unmarked. Overt pronoun subjects in Italian would then be the marked form, and therefore their interpretation would be dependent on the meaning of the unmarked null subject and as such require an additional reasoning step. Such an additional step

of reasoning about the speaker's choices when interpreting pronouns is argued to be effortful and to give rise to errors in child language (e.g., De Hoop & Kramer, 2006; Hendriks & Spenader, 2006).

Secondly, we will test the hypothesis that in languages in which there are fewer alternative forms, anaphoric subjects are processed differently from Italian. This would be the case for languages that have overt pronoun subjects, but no null subjects. We expect that in such non-null-subject languages the processing and interpretation of overt pronoun subjects does not depend on reasoning about alternative forms. We will investigate this by comparing the interpretation and processing of anaphoric subjects in Italian to the non-null-subject language Dutch. The data on Dutch are described in detail in Vogelzang et al. (2016), and will briefly be discussed below.

1.3 Comparison between anaphoric subjects in Italian and Dutch

In their study on anaphoric processing in Dutch, Vogelzang et al. (2016) compared the interpretation and processing of overt pronouns to that of full NPs, which unambiguously referred to the discourse topic. Overt pronoun subjects were less frequently interpreted as referring to the topic than unambiguous full NP subjects, indicating that although listeners have a very strong tendency to interpret an overt pronoun as referring to the discourse topic (93%), non-topical interpretations are allowed occasionally. Moreover, interpretation questions about Dutch overt pronouns were answered slower than questions about full NPs (Vogelzang et al., 2016).

In addition to the offline measures of final interpretation and response time, Vogelzang and colleagues used the online measure of pupil dilation during the auditory presentation of the

stimuli. Based on pupil dilation measures, Vogelzang and colleagues found that overt pronoun subjects require more cognitive effort to be processed than full NP subjects.

Given these results for Dutch, a language that only has overt pronouns, we will investigate how the availability of null subjects in Italian affects the processing of anaphoric subjects. For the correspondence between the referential preferences of anaphoric subjects in Italian and Dutch there are four logical options, which are illustrated in Figure 2. If the meaning of Italian null or overt pronouns corresponds directly to the meaning of Dutch overt pronouns, then one first possibility is that Italian *null* pronouns and their meanings correspond to Dutch *overt* pronouns and their meanings (option 1). Another option is that Italian *overt* pronouns and their meanings correspond to Dutch *overt* pronouns and their meanings, although a further category is present in Italian (option 2). Illustrations of these first two options are given in Example 3a and 3b, respectively (Italian sentences from Carminati, 2002, p.33):

(3)   a. Quando Mario$_i$ ha telefonato a Giovanni, $\varnothing_i$ aveva appena finito di mangiare. (Italian)

       Toen Mario$_i$ Giovanni belde, was hij$_i$ net klaar met eten. (Dutch)

       When Mario$_i$ called Giovanni, $\varnothing_i$ had just finished eating.

   b. Quando Mario ha telefonato a Giovanni$_i$, lui$_i$ aveva appena finito di mangiare. (Italian)

       Toen Mario Giovanni$_i$ belde, was hij$_i$ net klaar met eten. (Dutch)

       When Mario called Giovanni$_i$, he$_i$ had just finished eating.

It is, however, also conceivable that Italian and Dutch pronominal forms do not express exactly the same meanings. In this case, there is no direct correspondence between the meanings of Italian and Dutch pronominal forms, and thus one form is more specific in its meaning than another. One option is that Italian overt and null pronouns have a more specific meaning than

Dutch overt pronouns (option 3). In sentences 3a and b, for example, this would mean that the Italian pronouns have a strong preference to refer to either Mario or Giovanni, but the Dutch pronoun could refer to both and would thus be strongly ambiguous. An alternative option is that Italian overt and null pronouns have a less specific meaning than Dutch overt pronouns (option 4). This could be the case, for example, if Italian null pronouns in sentences like 3a were ambiguous, whereas Dutch overt pronouns always referred to Mario.

- Figure 2 here -

2 Current study

In this study, we examine the processing of anaphoric subjects in Italian. By means of pupil dilation measures, we examine the cost of online processing of different kinds of anaphoric subjects: non-pronominal full NPs such as *the hedgehog*, the overt pronouns *lui* 'he' and *lei* 'she', and null subjects. We compare the results of our Italian study to the results from the Dutch study by Vogelzang et al. (2016). Both experiments recorded responses, response times, and pupil dilation during a referent selection task.

Pupil dilation has been found to reflect cognitive processing effort (see, a.o., Beatty, 1982; Beatty & Lucero-Wagoner, 2000). Specifically, pupil dilation can serve as a measure of cognitive effort in language processing (e.g., Beatty & Lucero-Wagoner, 2000; Engelhardt, Ferreira, & Patsenko, 2010; Hyönä, Tommola, & Alaja, 1995; Just & Carpenter, 1993; Scheepers & Crocker, 2004; Schmidtke, 2014; Zellin, Pannekamp, Toepel, & Van der Meer, 2011) and can provide information about the processing of sentences, even without an explicit response from the participant, or provide signatures of cognitive processes that cannot be

derived from overt responses (e.g., Van Rijn, Dalenberg, Borst, & Sprenger, 2012; Wierda, Van Rijn, Taatgen, & Martens, 2012). This method has successfully been applied in research on processing effort during spoken language comprehension, specifically in studies on syntactic ambiguity resolution (Engelhardt et al., 2010), subject anaphora resolution (Vogelzang et al., 2016), and object anaphora resolution (Van Rij, 2012).

In the current study, increased cognitive effort is predicted to be reflected by inconsistent answers on the referent selection task, longer response times, and increased pupillary responses. We expect more cognitive effort to be needed in Italian to resolve marked overt pronoun subjects compared to unmarked null subjects. Unambiguous full NP subjects are expected to be easier to process than both types of pronouns, because, in contrast to ambiguous pronouns, no ambiguity resolution is needed for unambiguous NPs (cf. Vogelzang et al., 2016).

A comparison will be made between our findings on the interpretation of anaphoric subjects in Italian and earlier findings on Dutch. Our experiment on Italian uses translated versions of the stimuli of the Dutch study of Vogelzang et al. (2016), to allow for an optimal comparison between anaphoric interpretations in the two languages. Across the two experiments, similar participant populations and methods were used. Any differences between the Italian experiment and the experiment of Vogelzang and colleagues will be explicitly mentioned in the following sections.

3 Methods

3.1 Participants

Forty students from the University of Milano-Bicocca, Italy (12 men; mean age 23.5; SD =

5.8) participated in the experiment for course credits. All participants were monolingual native speakers of Italian. All participants had normal or corrected-to-normal vision and hearing.

3.2 Materials

In the experiment, participants heard short stories, each followed by a question about the second or final clause of the story. The experiment used translated versions of the Dutch stories used in Vogelzang et al. (2016). Due to time constraints, we could not use all 120 stories from the Vogelzang et al. experiment, but randomly selected 24 stories for removal, leaving 96 stories in the experiment. While the Dutch experiment investigated the interpretation of two types of anaphoric subjects (full NPs and overt pronouns), the Italian experiment investigates three types of anaphoric subjects (full NPs, overt pronouns, and null subjects). In addition, the Dutch experiment also investigated the interpretation of two types of anaphoric objects (reflexives and pronouns). In order to maximize parallelism between the current experiment on Italian and Vogelzang et al.'s earlier experiment on Dutch, the current experiment on Italian uses the same object manipulation.

A similar story structure was created for all experimental items. Each story featured two referents, which were displayed as pictures on the screen and remained there for the whole duration of the trial. All stories consisted of a succession of three clauses. The first clause was a simple sentence, in which the first referent was introduced as the grammatical subject. The second clause was part of a complex sentence in which the same referent was mentioned again as the subject, making it unambiguously the discourse topic. A second, non-subject, referent was introduced as well in this second clause. Thus, by the end of the second clause, two referents had been mentioned: one (the subject referent) being more prominent than the other (the non-subject referent). The third and final clause of each story started with the temporal

conjunction *mentre* 'while' (cf. Serratrice, 2007; Sorace & Filiaci, 2006), which introduces a subordinate clause without creating a causal bias. The subject of this third clause was one of three anaphoric subjects: a full NP explicitly referring to the topical subject (e.g., *il riccio* 'the hedgehog'), an overt pronoun (either *lui* 'he' or *lei* 'she'), or a null subject (Ø). In the third clause, we also manipulated the type of object, which was either the reflexive *si* ('himself/herself') or a clitic pronoun (*lo* 'him' or *la* 'her'). The gender of the pronouns always matched both referents in the story, so gender was not a discriminating cue. An example story is presented in Table 1.

Following each story, a referent selection question was asked to determine the participant's interpretation of the anaphoric subject in the final clause of the story. Questions corresponding to the story in Table 1 are shown in Table 2. In order to counterbalance the expected answers to the referent selection questions and thus reduce predictability, we also included additional questions asking about the interpretation of the object pronoun or about the non-topical referent in the second clause of the stories.

- Table 1 here -

- Table 2 here -

3.3 Design

The experiment used a within-subjects design with anaphoric subject as the critical condition (full NP, overt pronoun, and null pronoun). Additionally, anaphoric object (reflexive and pronoun) was manipulated as a condition. The experiment consisted of 96 items in total (16 critical items per subject-object combination), distributed over four blocks. Before the main

experiment, participants received a practice block of six stories. The complete experiment took approximately 60 minutes. The order of presentation of the stories was pseudo-randomized per participant.

3.4 Procedure

Following the same procedure as Vogelzang et al. (2016), participants were tested individually using an EyeLink 1000 (SR Research) eye tracker with head support to measure pupil dilation and a Microsoft Sidewinder gamepad to record responses. Participants heard the stories and questions through headphones while looking at a computer screen. They were instructed to listen carefully and to blink as little as possible during trials. To reduce the number of blinks during trials, participants were explicitly asked to blink after every trial.

Throughout each trial, the two referents in each story were presented as pictures on the screen; one on the left side and the other on the right side. The location of the topical referent was counterbalanced across trials. As the onset of the pictures would influence the pupillary measures, they appeared before the beginning of the story and remained visible until after the question had been answered. Participants could answer the referent selection question by selecting one of the referents on the screen through the left or right trigger button on the back of the gamepad. After the question was answered, a black frame was displayed around the selected referent. No other feedback was provided. Figure 3 shows the structure of presentation of the stimuli during a trial.

- Figure 3 here -

4 Results

The experiment recorded offline responses to the referent selection questions, the corresponding response times, and pupil dilation during sentence processing. Preprocessing was performed on the pupillary data. In the time window ranging from 1000 ms before the onset of the word following the critical anaphoric subject until the onset of the question, samples recorded during blinks were encoded as missing data. Trials that contained more than 25% missing data in this time window were excluded. The missing data of the remaining trials were interpolated. In total, 350 trials (9.1%) were excluded and 3490 trials remained for analysis. No further selection of the data, for example based on the correctness of the responses, was made, as we were interested in the processing of potentially ambiguous forms that do not have a clear-cut correct or incorrect interpretation. For the analyses of the responses and the response times, we selected the subset of the data in which the interpretation question was a question about the anaphoric subject, and not about the anaphoric object. A total of 1722 trials remained for analysis. The results of the offline responses will be presented first, followed by the response times (RTs) and the pupil dilation data.

4.1 Responses

4.1.1 Italian

Throughout the story exemplified in Table 1, the firstly introduced referent (e.g., *the hedgehog* in the story in Table 1) is the discourse topic on the basis of prominence, frequency, and subjecthood, whereas the other referent (e.g., *the mouse* in the story in Table 1) is a non-topical referent. If the answer to the referent selection question corresponded to the discourse topic, this was coded as a topic continuation. If, on the other hand, the answer to the question corresponded to the non-topical referent, this was coded as a topic shift. The coded responses

to the questions in the three anaphoric subject conditions for Italian are shown in Figure 4a (the responses split up per object condition are shown in Figure A1 in Appendix A). The responses are expressed as the percentage of responses indicating a topic continuation.

- Figure 4 (a&b) here -

The responses to the interpretation questions indicate that in Italian, unambiguous full NP subjects were consistently interpreted correctly as referring to the discourse topic (99%). As expected, null subjects were generally interpreted as referring to the discourse topic as well (86%). For overt pronoun subjects, there were both topic continuation interpretations (39%) and topic shift interpretations (61%).

To examine whether the interpretations of overt pronoun subjects differed from chance, a Welch two-sample t-test was performed. The results showed that the antecedent preference for overt pronoun subjects (61%) significantly deviated from 50%, and thus that overt pronouns show a preference towards a non-topic antecedent ($t(39) = 3.43$; $p < 0.01$).

The responses were analyzed using a binomial generalized linear mixed effect-based model. Treatment coding was used throughout the paper, which means that all comparisons were done relative to a baseline. In this model, a baseline of full NP subjects and reflexive objects was used. As in Vogelzang et al. (2016), we included anaphoric subject (full NP, overt pronoun, null pronoun), anaphoric object (reflexive, pronoun), and the interaction between subject and object as fixed effects. Following Baayen, Davidson, and Bates (2008), we assessed, by means of model comparisons, whether the inclusion of random factors and co-variates such as participant and item number, trial order, side on which the topic referent was displayed, gender

of the referents, gender and age of participant was warranted. The best model included both random intercepts for participants, random intercepts for items, and random slopes for object form within items.

The output of the model that best fits the data is reported in Table A1 in the Appendix. It shows that participants gave more topic continuation responses with full NP subjects than with overt pronoun subjects (β = -6.57; z = -6.40; p < 0.001) or null subjects (β = -4.24; z = -4.15; p < 0.001) in sentences with a reflexive object. In sentences with a full NP subject, no difference was found between reflexives objects and pronoun objects (β = -1.13; z = -1.10; p = .271); with both objects forms, responses were near ceiling level (100% topic continuation with reflexive objects, 98% topic continuation with pronoun objects, see also Figure A1 in the Appendix).

To be able to compare between null and overt pronoun subjects, we ran an additional model with a baseline of overt pronoun subjects and reflexive objects. The model included anaphoric subject, anaphoric object, and their interaction as fixed effects, random intercepts for participants and items, and random slopes for object form within items. The model revealed that participants selected more topic continuation responses with null pronoun subjects than with overt pronoun subjects (β = 2.34 z = 9.82; p < 0.001; for the full model results see Table A2 in the Appendix).

4.1.2 Comparison between Italian and Dutch

For all comparisons with Italian we used the subset of the Dutch data with the same stories that were used in the Italian study. The Dutch dataset in our comparison therefore consisted of 96 stories (of the 120 stories) and 40 participants. In this combined dataset, both versions of a story (items) received identical dummy-coded values for statistical analyses. The responses to

the questions in the two anaphoric subject conditions from the Dutch study are shown in Figure 4b. A generalized linear mixed-effects model with language, anaphoric subject, and anaphoric object as fixed (categorical) effects and random intercepts for participants and items with a baseline of Italian full NP subjects and Italian reflexive objects showed no differences in the interpretation of full NP subjects between Italian and Dutch (99% vs. 99%, $\beta = -0.17$; $z = -0.31$; $p = 0.76$). In order to compare Italian null subjects and Dutch overt pronoun subjects, an additional model was made including language, subject form, and object form as fixed effects, and random intercepts for participants and items with a baseline of Italian null pronoun subjects and Italian reflexive objects. The results showed that Dutch overt pronoun subjects (93% topic continuation) have a stronger preference for the topical antecedent than Italian null pronoun subjects (86% topic continuation; $\beta = 0.79$; $z = 3.21$; $p = 0.001$). The complete model results are listed in Appendix B (Table B1, B2).

4.2 Response times

4.2.1 Italian

The response times (RTs) to the interpretation questions were measured from the end of the question until the first button press. The mean RT per subject condition for Italian is shown in Figure 5a (the response times split up per object condition are shown in Figure A2 in the Appendix).

- Figure 5 (a&b) here -

The RTs were analyzed using a linear mixed effect-based model of the log-transformed RTs, with a baseline of full NP subjects and reflexive objects. Based on the analyses of Vogelzang

et al. (2016), we included anaphoric subject, anaphoric object, and the interaction between subject and object as fixed effects. Based on model comparisons (cf. Baayen et al., 2008), we also included random intercepts for participants and random slopes for anaphoric subjects within participants.

The model results (Table A3 in the Appendix) show that full NP subjects are responded to faster than overt pronoun subjects (β = 0.47; t(85) = 5.28; p < 0.001) and null subjects (β = 0.49; t(77) = 5.21; p < 0.001) in sentences with a reflexive object. In object position, a difference was found between reflexives and pronouns (β = 0.17; t(1614) = 2.26; p = 0.02), indicating that sentences with an NP subject were responded to slower when a pronoun object was used than when a reflexive object was used (see also Figure A2 in the Appendix).

To test for differences between null and overt pronoun subjects, we ran an additional model including anaphoric subject, anaphoric object, and their interaction as fixed effects, random intercepts for participants and random slopes for subject form within participants. Overt pronoun subjects and reflexive objects were the baseline. The model revealed no difference in RTs between overt and null pronoun subjects (β = 0.01; t(116) = 0.18; p = 0.86).

4.2.2 Comparison between Italian and Dutch

The RTs from the Dutch study are shown in Figure 5b. We ran a linear mixed-effects model with language, subject type, and object type as fixed effects and random intercepts for participants and items, and Italian full NP subjects and Italian reflexive objects as baseline. The model revealed no differences between the response times to Italian and Dutch full NP subjects (resp. 549 ms and 583 ms, β = 0.03; t(110) = 0.32; p = 0.751). An additional model was run with a baseline of Italian null pronoun subjects and Italian reflexive objects. This model

included language, subject, and object as fixed effects, and random intercepts for participants and items. The results showed no significant difference between the response times to Italian null pronoun subjects and Dutch overt pronoun subjects (resp. 909 ms vs. 726 ms, β = -0.103; $t(82)$ = -1.19; p = 0.237). A final model with language, subject, and object as fixed effects, random intercepts for participants and items, and Italian overt pronoun subjects and Italian reflexive objects as a baseline, showed no significant difference between the response times to Italian overt pronoun subjects and Dutch overt pronoun subjects (903 ms vs. 726 ms, β = -0.08; $t(82)$ = -0.94; p = 0.352). The output of these models are listed in Appendix B (Table B3-B5).

4.3 Pupil dilation data

Before further analysis, the pupil dilation data were down-sampled to a sampling rate of 100 Hz. For the analysis of the pupil dilation data, we considered the time window from 1000 ms before the onset of the word following the critical anaphoric subject in the third clause until the onset of the question. Because the time window did not include the referent selection questions, the full data set was used for the pupil dilation analysis.

4.3.1 Italian

The pupil dilation data for each anaphoric subject over the course of a trial, averaged over all trials, are plotted in Figure 6 (the pupil dilation data split up per object condition are shown in Figure A3 in the Appendix). The dilation is expressed as proportional change relative to a baseline window. This baseline was determined by first calculating the mean pupil dilation for each trial separately, from 100 ms before the onset of the subject until the onset of the subject (corresponding to the second half of the word *mentre,* 'while'). The samples from each trial were subsequently divided by their corresponding baseline in order to get the proportional pupil

dilation. A positive proportional pupil dilation indicates that the pupil size, interpreted as reflecting processing effort, increased during the test sentence.

The dilation data were aligned to the onset of the word that followed the anaphoric subject, which was the anaphoric object. This was done because the anaphoric subjects in the experiment varied in length (e.g., *il riccio* vs. *lui* vs. $\varnothing$). The mean onset of the anaphoric subjects in the plot therefore varies from -1000 ms for full NPs to -500 ms for overt pronouns, and to 0 ms for null pronouns (for which the onset of the subject was taken to be identical to the onset of the object). Thus, full NPs already start around 1000 ms before t=0 in Figure 6, and therefore the NP condition shows a rise in pupil dilation before t=0.

- Figure 6 here -

We used Generalized Additive Models (GAMs; Wood, 2006) to analyze the pupil dilation data. This allows for an analysis of the data over time (e.g., Boehm, Van Maanen, Forstmann, & Van Rijn, 2014; Van Rij, Hollebrandse, & Hendriks, 2016), which cannot be done with traditional linear mixed-effects models. Based on the analyses of Vogelzang et al. (2016), the model used anaphoric subject, anaphoric object, and the interaction between subject and object as fixed effects. Based on model comparisons, random slopes and intercepts for participants, random slopes and intercepts for items, an effect of trial order, and an effect of time over centered trials were included. The model used a baseline of Italian full NP subjects and Italian reflexive objects.

The results (see Table A5 in the Appendix) show, when examining the effect of the anaphoric subject, a difference in pupil dilation between sentences with overt pronoun and null subjects

compared to full NP subjects (resp. $F(1,17.90) = 63.43$, $p < 0.001$ and $F(1,17.12) = 22.00$, $p < 0.001$). Figure 6 shows that from around 1000 ms after the onset of the word following the subject, both the condition with an overt pronoun subject and with a null subject show more dilation than full NP subjects. From 2000 ms on, the pupil dilation stabilizes, with overt pronouns eliciting the most pupil dilation, full NPs eliciting the least pupil dilation, and null pronouns eliciting a pupil dilation in between overt pronouns and full NPs. An additional model with a baseline of null subjects and reflexive objects (see Table A6 in the Appendix) confirms that the pupil dilation curve for sentences with overt pronouns is different from the curve with null pronouns ($F(1,16.92) = 25.88$, $p < 0.001$; see Table A6 in the Appendix) in sentences with a reflexive object. Recall that the offsets of the different anaphoric subjects vary; this explains why in Figure 6 the dilation in sentences with overt pronouns starts rising earlier than with null pronouns.

When examining the different object forms (see the average pupil dilation data per object condition in Figure A4 in the Appendix), the results of the model show that when combined with full NP subjects, pronoun objects influence pupil dilation compared to reflexive objects (see Table A5; $F(1,17.21) = 4.25$, $p < .001$). However, the model fit of this effect (see Appendix A, Figure A5) does not show a clear difference in pupil dilation for these two object forms. A stronger effect is shown in the interactions between pronoun objects and null subjects ($F(2,16.71) = 9.21$, $p < .001$) and between pronoun objects and overt pronoun subjects ($F(2,15.11) = 6.75$, $p < .001$). Figure A3 and the model fits in Figure A6-7 in Appendix A show that, in these conditions, pronoun objects elicit more pupil dilation than reflexive objects.

4.3.2 Comparison between Italian and Dutch

In this section, we compare the Italian results to the Dutch results. Figure 7 presents the pupil

dilation data from the Dutch study of Vogelzang et al. (2016). The difference in amplitude between the Italian pupil dilation data in Figure 6 (max. amplitude around 0.09) and the Dutch pupil dilation data in Figure 7 (max. amplitude around 0.03) is clear, and is most likely caused by differences in lighting at the different test locations[4]. Pupil dilation is a measure that is very sensitive to influences that are unrelated to the experimental conditions, such as lighting conditions in the lab, the participant's attention and possibly language-specific influences, so it is not possible to compare the proportional dilation of the two experiments directly. Even though practically unfeasible, the results in the two languages would have been better comparable if the participants were tested in the same lab. However, by testing Italian speakers in the Netherlands or vice versa, additional factors such as testing in a first versus foreign language context emerge that might influence the results. Therefore, we chose to test Italians in Italy and report on a comparison between the *relative* effects found in the two experiments. Yet, Appendix C presents a small follow-up study to investigate whether Dutch and Italian participants' pupil dilation amplitude is more similar when tested in the same lab.

In Dutch, it was found that sentences with overt pronoun subjects elicit more pupil dilation than sentences with full NP subjects. This is in line with the Italian results, where sentences with overt pronoun subjects as well as sentences with null subjects elicit more pupil dilation than sentences with full NP subjects.

- Figure 7 here -

However, there are also differences between the Italian and the Dutch results. In Italian, the effects of processing the clause are already visible before the onset of the word following the anaphoric subject (in the case of an NP) or right after the onset of the word following the

22

anaphoric subject (in the case of an overt pronoun). In Dutch, the rise in pupil dilation in sentences with full NP subjects or overt pronoun subjects starts much later, around 1000 ms.

We assessed whether the increase in pupil dilation after the onset of the word following the subject differed as a function of language. Note that this analysis is post-hoc, as we did not expect to see any effects in slope based on theoretical considerations. To eliminate the difference in maximum amplitude between the Italian and Dutch datasets, we normalized all pupil dilation data to the mean of the amplitude between the offset of the sentence and the onset of the question as assessed in the NP condition. We calculated regression slopes of the pupillary response for the first 1000 ms after the onset of the word following the subject for each participant and for each condition separately for both Italian and Dutch. A post-hoc linear mixed effect-based model showed that the average slope of the Italian data is steeper than the average slope of the Dutch data (β = -0.41; t = -3.71; $p < 0.001$). The complete model results are listed in Appendix B (Table B6).

5 Discussion: A cross-linguistic comparison

In this paper, we described a referent selection study that investigated the interpretation and processing of anaphoric subjects in Italian, a null-subject language. We further compared the results of this study to results from an experiment in Dutch, a non-null-subject language. Our referent selection task measured interpretation, response time, and pupil dilation for sentences containing anaphoric subjects. The results of the Italian study are discussed first, and are then compared to the results of the Dutch study of Vogelzang et al. (2016).

5.1 Italian

23

We hypothesized that overt pronoun subjects would be the marked form in Italian, and would require the listener to take an additional step of reasoning about alternative forms. This additional reasoning step was predicted to be reflected by less consistent answers on the referent selection task, longer response times, and increased pupillary responses. Full NP subjects were expected to be easier to process than both overt and null pronoun subjects.

As predicted, the results show effects of anaphoric subject form, with unambiguous full NPs being interpreted more often as referring to the discourse topic than potentially ambiguous pronouns.

In the introduction, we suggested that in Italian null subjects are the unmarked form, and overt pronoun subjects are the marked form (cf. Horn's division of pragmatic labor, 1984). Our results indeed show a division of pragmatic labor, because the different anaphoric subjects have different meanings. In line with Carminati (2002), null subjects were generally interpreted as signaling a topic continuation, while overt pronoun subjects were most often interpreted as signaling a topic shift in our experiment. In other words, null subjects tend to refer to referents with a high prominence, and overt pronoun subjects tend to refer to referents with a lower prominence. This finding is similar to findings in other studies on Italian (e.g., Sorace & Filiaci, 2006) and Spanish (e.g., Alonso-Ovalle, Fernández-Solera, Frazier, & Clifton, 2002), although exact percentages may differ based on the task.

However, the results also show that the two Italian subject pronouns do not always have clearly distinct meanings; null subjects refer to the discourse topic 86% of the time and overt pronoun subjects refer to the same topic 39% of the time. Thus, at least part of the time these subject forms are interpreted as referring to the same entity. This indicates that Horn's division of

pragmatic labor with respect to anaphoric subjects in Italian is a gradient rather than absolute distinction.

When looking at the response times and the pupil dilation measures, full NP subjects are responded to faster, and elicit less pupil dilation than potentially ambiguous pronouns. These results are in line with our predictions about the processing of full NPs compared to pronouns, and indicate that processing ambiguous pronouns is more effortful than processing unambiguous full NPs. Thus, pupil dilation is a viable measure of processing effort.

The difference in processing overt versus null pronoun subjects was examined through response times and pupil dilation. Although we do not find any differences between the two pronominal forms in the response times, we do so in the pupil dilation results. Overt pronouns elicit more pupil dilation than null pronouns, indicating that processing overt pronoun subjects is more effortful than processing null subjects. The differences in processing effort that are reflected in pupil size during the sentence were no longer present or measurable after the referent selection question. Thus, these results show that the measure of pupil dilation is more sensitive than the measure of response times as a reflection of online processing difficulty.

Overall, the Italian results show differences in interpretation as well as online processing between the different anaphoric subjects, with more specific anaphoric subjects being easier to process than less specific anaphoric subjects. We argue that the difficulties in the interpretation and processing of marked overt pronoun subjects stem from the execution of a costly additional reasoning step, which is necessary to reason about alternative forms (cf. Blutner, 2000; Hendriks et al., 2010; Horn, 1984).

5.2 Comparison between Italian and Dutch

Unambiguous full NPs are interpreted quickly and correctly in both Italian and Dutch. Regarding Dutch overt pronouns, Vogelzang et al. (2016) showed they preferably refer to a topical antecedent. This preference was also found for Italian null pronouns in the present study. Although we did not find significant differences in response times between Italian and Dutch pronouns, the antecedent preferences of Dutch overt pronouns were stronger than those of Italian null pronouns.

In the introduction, we presented four logical options regarding the relation between Italian and Dutch anaphoric subjects. The finding that Dutch pronoun subjects indicate a topic continuation more often than Italian null subjects provides evidence for Option 4 (see Figure 2). As discussed in the introduction, the difference in meaning between anaphoric subjects in Italian and Dutch may be related to the availability of alternative anaphoric subject forms in these languages[5]. When a language offers a choice between a marked and an unmarked form (cf. Horn, 1984), as with pronouns in Italian, as opposed to the availability of a single form, as with pronouns in Dutch, the choice of a specific form provides additional information about the intended referent. This may restrict the possible meanings for the form (cf. Blutner, 2000; Hendriks et al., 2010; Kiparsky, 1982). This additional information has to be considered when resolving the referential ambiguity of the anaphoric subject and leads to a specific meaning for the pronoun. Additionally, because in Italian an overt pronoun subject tends to take a non-subject antecedent, using a pronoun to refer to a non-subject antecedent is more common in Italian than in Dutch. Therefore, the difference in interpretation between Italian null subjects and Dutch overt pronoun subjects may be influenced by the possibilities and frequency of referring to non-subjects, reflecting linguistic differences between the two languages.

In the pupil dilation data analyses, the effect of pronouns being costlier to process than full NPs is evident in both languages. An unexpected finding was the difference in onset of the pupillary effects over time. The pupillary responses in Dutch were delayed compared to Italian. Although such an effect was not included in our initial hypotheses, the earlier effects for Italian could indicate that Italian anaphoric subjects are more informative than Dutch anaphoric subjects, and therefore elicit an earlier pupillary increase. Further research on the timing of pupillary responses (such as the work of Wierda, Van Rijn, Taatgen, & Martens, 2012, using a temporal attention task) is necessary to be able to fully interpret delayed effects.

5.3 General issues

In this paper, we examined the processing of different types of anaphoric subjects (null, overt, and full NPs), and compared the findings in two languages with a different pronominal system, one that allows for null pronoun subjects (Italian), and one that does not (Dutch). We found that anaphoric subjects in Italian and Dutch are interpreted differently, and, as we said, we conjectured that this may be related to the difference in availability of alternative anaphoric subject forms in these languages.

However, the results can be generalized beyond the specific scope of this study. Three general questions are discussed below: 1) what are the advantages and disadvantages of using pupil dilation as a measure of processing effort?, 2) how can the results be extended to the processing of expressions other than anaphoric subjects?, and 3) how should words that have similar but slightly different meanings be categorized?

Firstly, the study showed the value of examining processing effort and ambiguity resolution during sentence processing by means of pupillometry. Whereas the response times to

interpretation questions did not differentiate between Italian overt pronoun subjects and Italian null subjects, the online measure of pupil dilation showed a clear effect. Additionally, the pupil dilation data revealed differences in the timing of pronoun processing across languages. Thus, pupillary responses are a sensitive measure of online processing effort that can reveal effects that offline measures do not demonstrate.

Nevertheless, the method of pupil dilation measurement also has some disadvantages. Firstly, it is a relatively slow measure, with peak latencies having been reported up to 1.3s after the critical stimulus (Just & Carpenter, 1993). In the current study, peak latencies were even delayed to 2-3s, although differences between conditions already appear earlier. Thus, experimental designs should allow for sufficient time between the critical word and the response. Secondly, pupil size can be influenced by cues and factors other than the presented item, such as attention and lighting conditions. This made it impossible in our study to directly compare the absolute pupil dilation values of the Dutch and the Italian studies. For this, other measures would have been more appropriate, such as the Index of Cognitive Activity (Marshall, 2002), which is a measure of pupil size that is not confounded by lighting conditions, or EEG, which is a fast measure of processing that has also been applied to measure cognitive load during sentence processing (e.g., Marsella et al., 2017). However, these other measures have their own disadvantages.

In this paper, it was assumed that increased pupillary responses reflect cognitive effort of language processing, as has been suggested in the literature (e.g., Beatty & Lucero-Wagoner, 2000; Engelhardt, Ferreira, & Patsenko, 2010). However, other factors may also have influenced pupil size, such as (implicitly or explicitly) deciding to move the eyes towards a referent (cf. Katidioti, Borst, & Taatgen, 2014). In our study, all these potential influences are

expected to result in a larger pupil size when a referent is more difficult to process. So, whether the measured pupil size reflects referent processing directly or reflects processes related to referent processing, the differences in pupil size found in the experiment can still be interpreted as reflecting cognitive effort. However, more research on pupillary responses is needed to better understand the many factors influencing pupil dilation.

Regarding the second question posed above, namely how the results on anaphoric subjects can be extended to other expressions, the finding that overt pronoun subjects in Italian are more effortful to process than null subjects may be extended to the processing of other pairs of expressions that have seemingly similar meanings, such as anaphoric objects. Italian distinguishes between clitics (e.g., *lo* 'him') and strong or stressed forms (e.g., *lui* 'him'). On the basis of the results presented in this paper, we would predict that anaphoric objects also show a division of labor between an unmarked and a marked form, resulting in differences in processing effort.

This prediction of processing differences for pairs of expressions that have similar meanings also has possible implications for the acquisition and use of the mental lexicon. When multiple expressions are available that seemingly refer to the same entity, like 'coach' and 'bus', processing could be influenced by how specific the expression is. In other words, expressions that are assumed to be synonyms in certain contexts may not be equally easy to process. To investigate this issue, many studies that have examined the processing and interpretation of synonymous expressions could be repeated, now also measuring processing difficulty by means of pupil dilation.

As for the final question posed above, the study shows that seemingly similar forms in different

languages can have different meanings. Words with a similar phonology or orthography in different languages, but with different meanings, have previously been named *false friends* (Aronoff & Rees-Miller, 2002). Examples of a false friend are the Italian 'attuale', which in English means 'present', but not 'actual' (Ferguson, 1994), and the Italian 'caldo', which means 'hot', not the phonologically and orthographically more similar 'cold'. We speculate that the anaphoric subjects used in our cross-linguistic comparison illustrate a different category of cross-linguistic 'friends', namely that of syntactically similar forms with different meanings, which we will call *fake friends*. Fake friends, unlike false friends, do not necessarily have a similar phonology or orthography. Nonetheless, they might lead to confusion because they belong to the same syntactic category and have the same syntactic function, and this, in turn, might lead to the wrong suggestion that the two forms have the exact same meaning, while, in fact, their meanings differ in subtle ways. In particular, the two forms interact differently with their linguistic context. Thus, whereas false friends are words that have no semantic overlap, fake friends have a large, albeit incomplete, semantic overlap. We suggest that Dutch overt pronoun subjects and Italian overt and null pronoun subjects do not adhere to the classical definition of false friends, but can be seen as examples from the distinct category of fake friends.

To conclude, we investigated how the availability of alternative referential forms in a language affects the processing and interpretation of a referential form. The results show that different anaphoric subjects are processed and interpreted differently in Italian. Moreover, different anaphoric subjects are processed and interpreted differently in the null-subject language Italian compared to the non-null-subject language Dutch. We argue that the processing of a form is affected by the availability of alternative forms for the same meaning in the language. This

assumes that an additional reasoning step is needed for the processing of forms for which competing alternatives are available.

Notes

[1] Different terms have been used to describe a referent's prominence in the discourse, such as accessibility (Ariel, 1990), givenness (Gundel et al., 1993), and topicality (Givón, 1983). We will use the notion of prominence as an operationalization of a multitude of factors that influence how accessible, given, or topical a referent is.

[2] There are additional anaphoric subjects that were not taken into account in this paper. For example, Italian and Dutch allow for a distinction between strong and weak pronouns and between full pronouns and clitics in some cases. The presented diagram only contains the anaphoric subjects that were used in the experimental studies and is therefore a simplification.

[3] The complete list of stimuli used in the experiment can be found on http://let.webhosting.rug.nl/~vogelzang/experiments.html or by contacting one of the authors.

[4] One reason for the difference in amplitude in the different studies could therefore be the lighting conditions in the labs. That is, due to different lightning conditions, the baseline pupil could have been closer to a minimal or maximal level of dilation. Therefore, we might have observed floor or ceiling effects. Our results suggest that that it was darker in the lab in Groningen in which the Dutch participants of Vogelzang et al. (2016) were tested than in the lab in Milan in which our Italian participants were tested. Regrettably, no reliable data is available on this difference. However, our assumption is that the pupils of the participants of Vogelzang et al. were already more dilated and thus expanded less.

[5] When comparing anaphoric subjects in different languages, it is unavoidable to test these

subjects in combination with other aspects of the languages. Thus, there are other factors that could have influenced the observed processing differences, such as pragmatic, syntactic, morphological, and lexical factors. The experiments were kept as similar as possible in design and materials in order to allow for a direct comparison between Italian and Dutch, but nevertheless care must be taken when interpreting the observed differences.

Acknowledgements

References

Alonso-Ovalle, L., Fernández-solera, S., Frazier, L., & Clifton, C. (2002). Null vs. overt pronouns and the topic-focus articulation in Spanish. *Rivista Di Linguistica*, *14*, 151–170.

Ariel, M. (1990). *Accessing noun-phrase antecedents*. London: Routledge.

Aronoff, M., & Rees-Miller, J. (2002). *The Handbook of Linguistics*. Malden, MA: Blackwell Publishing. http://doi.org/10.1002/9780470756409

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. http://doi.org/10.1016/j.jml.2007.12.005

Beatty, J. (1982). Task-Evoked Pupillary Responses, Processing Load, and the Structure of Processing Resources. *Psychological Bulletin*, *91*(2), 276–292.

Beatty, J., & Lucero-Wagoner, B. (2000). The pupillary system. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of psychophysiology* (pp. 142–162). New York: Cambridge University Press.

Belletti, A., & Guasti, M. T. (2015). *The Acquisition of Italian: Morphosyntax and its interfaces in different modes of acquisition*. Amsterdam/Philadelphia: John Benjamins Publishing Company. http://doi.org/10.1075/lald.57

Blutner, R. (2000). Some aspects of optimality in natural language interpretation. *Journal of Semantics*, *17*(3), 189–216. http://doi.org/10.1093/jos/17.3.189

Boehm, U., Van Maanen, L., Forstmann, B., & Van Rijn, H. (2014). Trial-by-trial fluctuations in CNV amplitude reflect anticipatory adjustment of response caution. *Neuroimage*, *96*, 95–105.

Carminati, M. N. (2002). *The processing of Italian subject pronouns* (Doctoral dissertation).

Amherst, MA: University of Massachusetts.

Chafe, W. L. (1994). *Discourse, consciousness, and time.* Chicago: Chicago University Press.

De Hoop, H., & Kramer, I. (2006). Children's optimal interpretations of indefinite subjects and objects. *Language Acquisition*, *13*(2), 103–123. http://doi.org/10.1207/s15327817la1302_4

Engelhardt, P. E., Ferreira, F., & Patsenko, E. G. (2010). Pupillometry reveals processing load during spoken language comprehension. *The Quarterly Journal of Experimental Psychology*, *63*(4), 639–645. http://doi.org/10.1080/17470210903469864

Ferguson, R. (1994). *Italian False Friends*. Toronto: University of Toronto Press.

Filiaci, F. (2010). *Anaphoric Preferences of Null and Overt Subjects in Italian and Spanish: a Cross-linguistic Comparison* (Doctoral dissertation). Edinburgh: University of Edinburgh.

Givón, T. (1983). Topic continuity in discourse: An introduction. In T. Givón (Ed.), *Topic Continuity in Discourse: A Quantitative Cross-language Study* (pp. 3–41). Amsterdam: John Benjamins Publishing Company. http://doi.org/10.1075/tsl.3

Gordon, P. C., Grosz, B. J., & Gilliom, L. A. (1993). Pronouns, names, and the centering of attention in discourse. *Cognitive Science*, *17*(3), 311–347. http://doi.org/doi:10.1207/s15516709cog1703_1

Greenberg, J. H. (1966). *Language Universals: with Special Reference to Feature Hierarchies*. The Hague, Netherlands: Mouton.

Grice, H. P. (1975). Logic and Conversation. In P. Cole & J. Morgan (Eds.), *Syntax and semantics: Speech acts* (Vol. 3, pp. 41–58). New York: Academic Press.

Gundel, J. K., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language*, *69*(2), 274–307. http://doi.org/10.2307/416535

Hendriks, P., De Hoop, H., Kramer, I., De Swart, H., & Zwarts, J. (2010). *Conflicts in Interpretation*. London: Equinox Publishing.

Hendriks, P., & Spenader, J. (2006). When Production Precedes Comprehension: An Optimization Approach to the Acquisition of Pronouns. *Language Acquisition*, *13*(4), 319–348. http://doi.org/10.1207/s15327817la1304_3

Horn, L. R. (1984). Toward a new taxonomy for pragmatic inference: Q-based and R-based implicature. In D. Schiffrin (Ed.), *Meaning, Form and Use in Context (GURT '84)* (pp. 11–42). Washington: Georgetown University Press.

Hyönä, J., Tommola, J., & Alaja, A. M. (1995). Pupil dilation as a measure of processing load in simultaneous interpretation and other language tasks. *The Quarterly Journal of Experimental Psychology*, *48*(3), 598–612.

Just, M. A., & Carpenter, P. A. (1993). The intensity dimension of thought: Pupillometric indices of sentence processing. *Canadian Journal of Experimental Psychology*, *47*(2), 310–339.

Katidioti, I., Borst, J. P., & Taatgen, N. A. (2014). What happens when we switch tasks: Pupil dilation in multitasking. *Journal of Experimental Psychology: Applied*, *20*(4), 380–396. http://doi.org/10.1037/xap0000031

Kiparsky, P. (1982). Word-formation and the lexicon. In F. Ingemann (Ed.), *Proceedings of the Mid-America Linguistics Conference* (pp. 3–29). Lawrence, Kansas.

Marsella, P., Scorpecci, A., Cartocci, G., Giannantonio, S., Maglione, A. G., Venuti, I., … Babiloni, F. (2017). EEG activity as an objective measure of cognitive load during effortful listening: A study on pediatric subjects with bilateral, asymmetric sensorineural hearing loss. *International Journal of Pediatric Otorhinolaryngology*, *99*, 1–7. http://doi.org/10.1016/j.ijporl.2017.05.006

Marshall, S. P. (2002). The Index of Cognitive Activity: measuring cognitive workload. In

Proceedings of the IEEE 7th Conference on Human Factors and Power Plants. IEEE.
http://doi.org/10.1109/HFPP.2002.1042860

McCawley, J. D. (1978). Conversational implicature and the lexicon. In P. Cole (Ed.), *Syntax and Semantics 9: Pragmatics* (pp. 245–259). New York: Academic Press.

Scheepers, C., & Crocker, M. W. (2004). Constituent order priming from reading to listening: A visual-world study. In M. Carreiras & C. Clifton (Eds.), *The On-line Study of Sentence Comprehension: Eyetracking, ERPs, and Beyond* (pp. 167–186). New York: Psychology Press.

Schmidtke, J. (2014). Second language experience modulates word retrieval effort in bilinguals: Evidence from pupillometry. *Frontiers in Psychology*, *5*, 137. http://doi.org/10.3389/fpsyg.2014.00137

Serratrice, L. (2007). Cross-linguistic influence in the interpretation of anaphoric and cataphoric pronouns in English–Italian bilingual children. *Bilingualism: Language and Cognition*, *10*(3), 225–238. http://doi.org/10.1017/S1366728907003045

Sorace, A., & Filiaci, F. (2006). Anaphora resolution in near-native speakers of Italian. *Second Language Research*, *22*(3), 339–368. http://doi.org/10.1191/0267658306sr271oa

Tsimpli, I., Sorace, A., Heycock, C., & Filiaci, F. (2004). First language attrition and syntactic subjects: A study of Greek and Italian near-native speakers of English. *International Journal of Bilingualism*, *8*(3), 257–277. http://doi.org/10.1177/13670069040080030601

Van Rij, J. (2012). *Pronoun processing: Computational, behavioral, and psychophysiological studies in children and adults* (Doctoral dissertation). Groningen, the Netherlands: University of Groningen.

Van Rij, J., Hollebrandse, B., & Hendriks, P. (2016). Children's eye gaze reveals their use of discourse context in object pronoun resolution. *Empirical Perspectives on Anaphora*

Resolution, (277), 267–293.

Van Rijn, H., Dalenberg, J. R., Borst, J. P., & Sprenger, S. A. (2012). Pupil Dilation Co-
Varies with Memory Strength of Individual Traces in a Delayed Response Paired-
Associate Task. *PLoS ONE*, *7*(12). http://doi.org/10.1371/journal.pone.0051134

Vogelzang, M., Hendriks, P., & Van Rijn, H. (2016). Pupillary responses reflect ambiguity
resolution in pronoun processing. *Language, Cognition and Neuroscience*, *31*(7), 876–
885. http://doi.org/10.1080/23273798.2016.1155718

Wierda, S. M., Van Rijn, H., Taatgen, N. A., & Martens, S. (2012). Pupil dilation
deconvolution reveals the dynamics of attention at high temporal resolution.
Proceedings of the National Academy of Sciences of the United States of America,
109(22), 8456–8460. http://doi.org/10.1073/pnas.1201858109

Wood, S. N. (2006). *Generalized Additive Models: an introduction with R*. Boca Raton, FL:
Chapman and Hall/CRC Press. http://doi.org/10.18637/jss.v016.b03

Zellin, M., Pannekamp, A., Toepel, U., & Van der Meer, E. (2011). In the eye of the listener:
pupil dilation elucidates discourse processing. *International Journal of
Psychophysiology*, *81*(3), 133–141. http://doi.org/10.1016/j.ijpsycho.2011.05.009

Tables and Figures

Subject Form		Clause				
		Il riccio	va con l'aereo in Inghilterra.			
		De egel	gaat met het vliegtuig naar Engeland.			
		The hedgehog	*goes with the airplane to England.*			
		In precedenza	il riccio ha chiesto	al topo	che ora fosse in Inghilterra,	
		Eerder	vroeg de egel	aan de muis	hoe laat het was in Engeland,	
		Earlier	*the hedgehog asked*	*to the mouse*	*what time it was in England,*	
NP	■	mentre	**il riccio**	**si**	affrettava	verso l'aereo.
	■	terwijl	**de egel**	**zich**	haastte	naar het vliegtuig.
		while	***the hedgehog***	***himself***	*hurried*	*to the airplane.*
NP	■	mentre	**il riccio**	**lo**	seguiva	verso l'aereo.
	■	terwijl	**de egel**	**hem**	volgde	naar het vliegtuig.
		while	***the hedgehog***	***him***	*followed*	*to the airplane.*
overt	■	mentre	**lui**	**si**	affrettava	verso l'aereo.
	■	terwijl	**hij**	**zich**	haastte	naar het vliegtuig.
		while	***he***	***himself***	*hurried*	*to the airplane.*
overt	■	mentre	**lui**	**lo**	seguiva	verso l'aereo.
	■	terwijl	**hij**	**hem**	volgde	naar het vliegtuig.
		while	***he***	***him***	*followed*	*to the airplane.*
null	■	mentre	**Ø**	**si**	affrettava	verso l'aereo.
		while	***Ø***	***himself***	*hurried*	*to the airplane.*

| null		mentre	*Ø*	**lo**	seguiva	verso l'aereo.
		while	*Ø*	***him***	*followed*	*to the airplane.*

Table 1. Example of an experimental story in Italian and Dutch (with English glosses in italics)[3]. All six possible final clauses of the story are listed, sorted by anaphoric subject. The anaphoric subjects investigated in Italian are NP (full NP subject), overt (overt pronoun subject), and null (null subject). In this story *the hedgehog* is the discourse topic, and *the mouse* is a non-topical referent. Note that the subordinate clause introduced by *mentre* 'while' has word order SOV, both in Italian and in Dutch.

Object form	Object	Question	Topic continuation answer	Topic shift answer
reflexive	si	Chi si affrettava?	Il riccio	Il topo
	zich	Wie haastte zich?	De egel	De muis
	himself	*Who himself hurried?*	*The hedgehog*	*The mouse*
pronoun	lo	Chi seguiva qualcuno?	Il riccio	Il topo
	hem	Wie volgde iemand?	De egel	De muis
	him	*Who followed someone?*	*The hedgehog*	*The mouse*

Table 2. Questions following the story in Table 1 in Italian, together with Dutch versions and English glosses. For the different anaphoric subjects, the same questions could be used. For the different anaphoric objects, different questions were used. Following the story in Table 1, the answer *the hedgehog* is coded as a topic continuation, and the answer *the mouse* is coded as a topic shift.
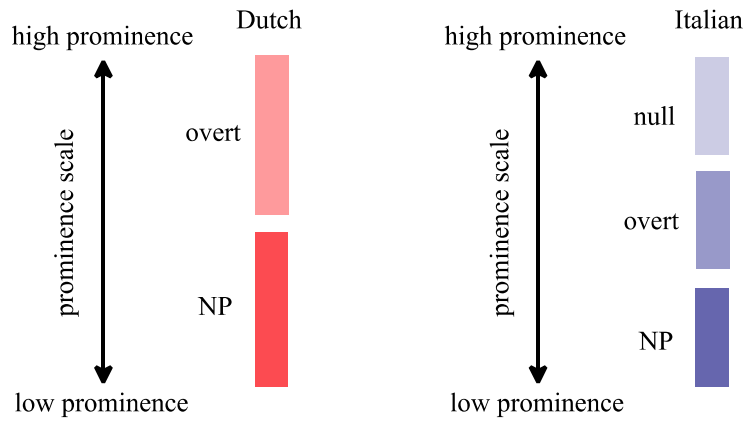
Figure 1. A prominence scale indicating that anaphoric subjects that are linguistically more constrained typically refer to referents with a low prominence, and that anaphoric subjects that are less constrained, like pronouns, typically refer to referents with a high prominence (NP: full NP subject, overt: overt pronoun subject, null: null subject). In the null-subject language Italian, an additional anaphoric subject is available compared to Dutch, namely a null pronoun[2].
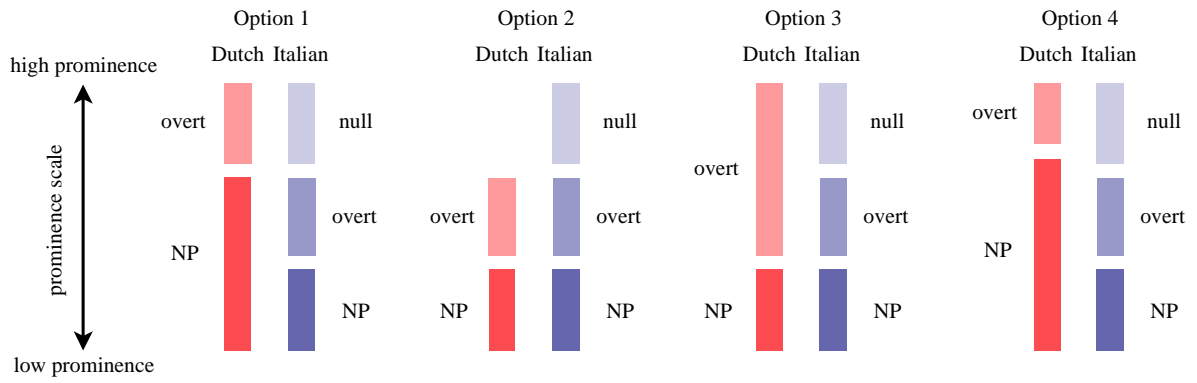
Figure 2. The four logical options for the correspondence between the forms of anaphoric subjects in Dutch and Italian (NP: full NP subject, overt: overt pronoun subject, null: null subject) along a prominence scale ranging from highly prominent referents (up) to low prominent referents (down).
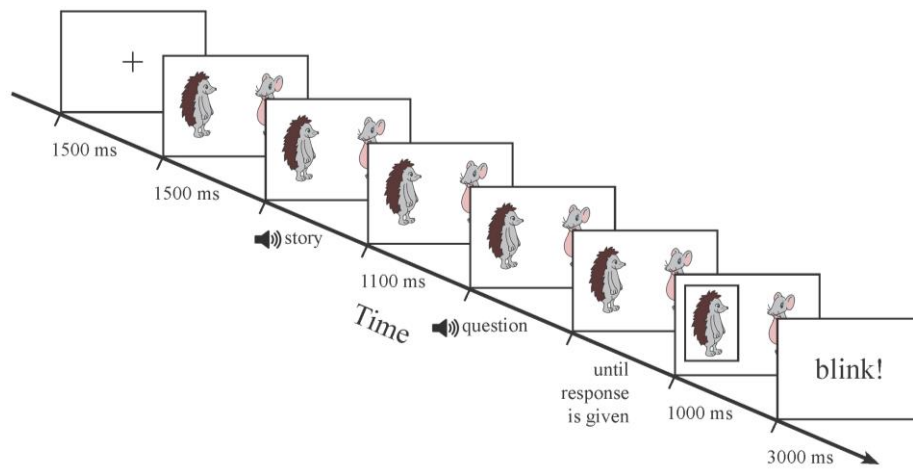
Figure 3. The structure of presentation of the stimuli during one trial. The story started 1500 ms after presentation of the referents. 1100 ms after the story, the question was presented. The participant could respond by selecting one of two referents on the screen. Figure taken from Vogelzang et al. (2016).
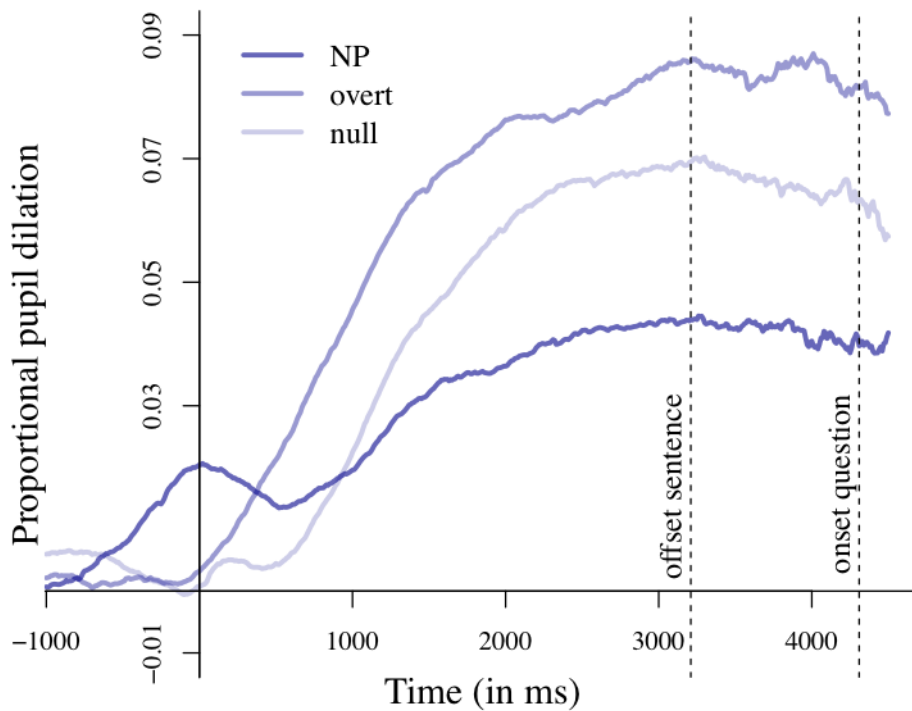
a: Italian   b: Dutch
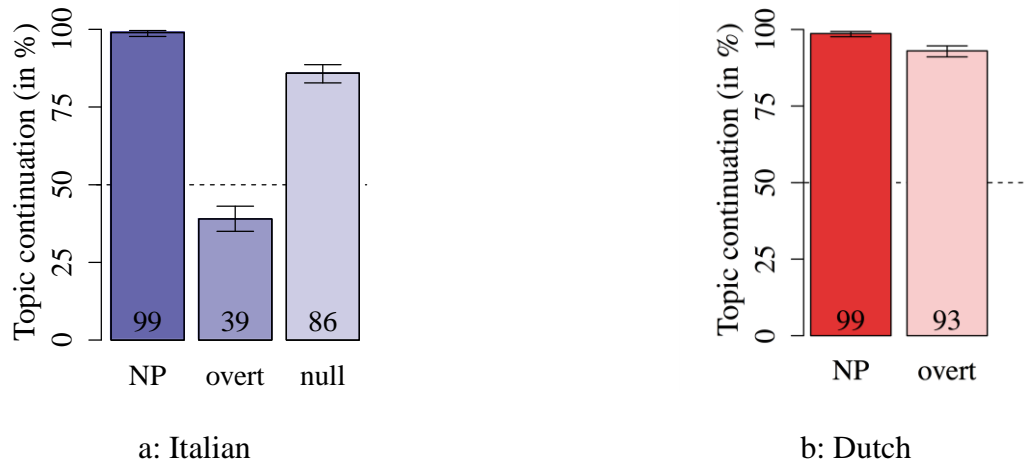
Figure 4. The mean percentage of responses indicating a topic continuation for the three anaphoric subjects in Italian (NP: full NP subject, overt: overt pronoun subject, null: null subject) are presented in Figure 4a. The mean percentage of responses indicating a topic continuation for the two anaphoric subjects in Dutch (NP: full NP subject, overt: overt pronoun subject) are presented in Figure 4b. Error bars are derived from logistic analysis.

a: Italian

b: Dutch
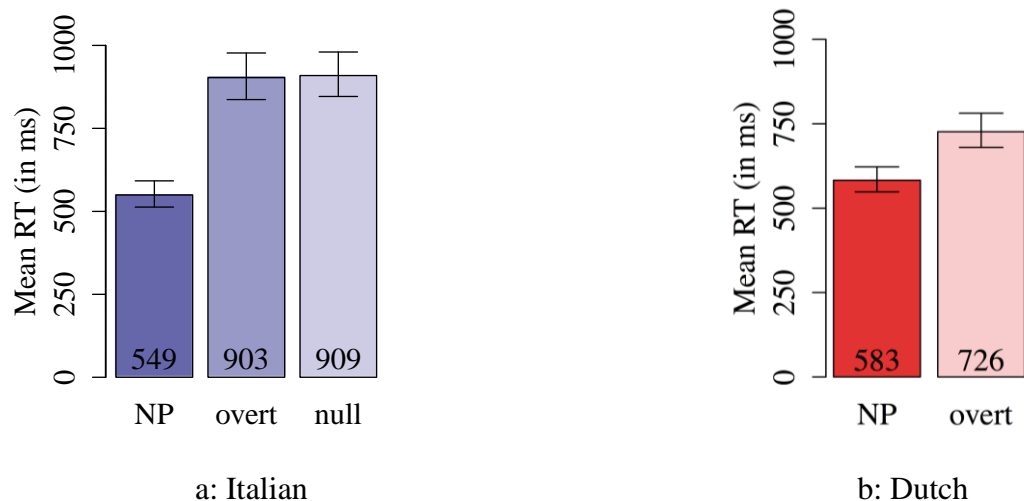
Figure 5. The mean response times for the interpretation questions regarding the three anaphoric subjects in Italian (NP: full NP subject, overt: overt pronoun subject, null: null subject) are presented in Figure 5a. The mean response times for the interpretation questions regarding the two anaphoric subjects in Dutch (NP: full NP subject, overt: overt pronoun subject) are presented in Figure 5b. Error bars are based on a 95% two-sided nonparametric confidence interval estimated through bootstrapping.
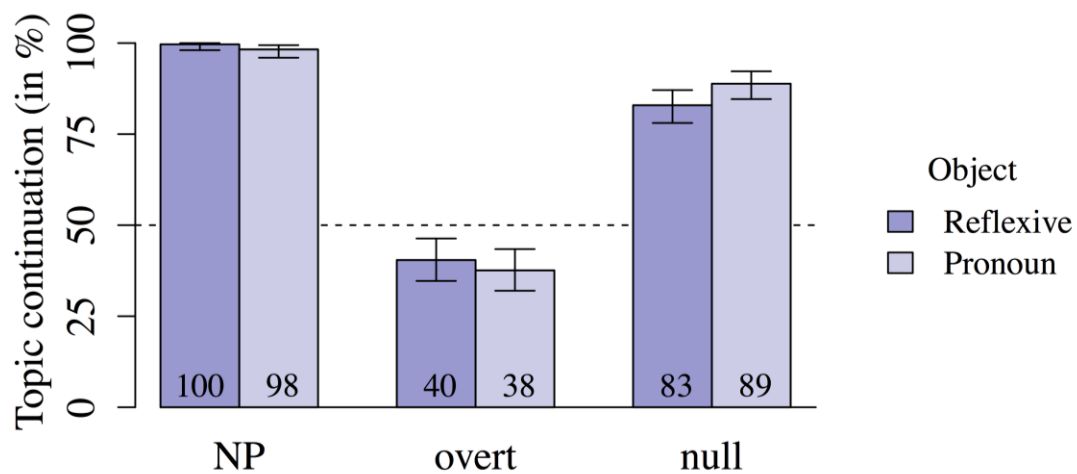
Figure 6. The mean proportional pupil dilation over time for the three anaphoric subjects in Italian (NP: full NP subject, overt: overt pronoun subject, null: null subject). Time is aligned to the onset of the word following the subject. Vertical dotted lines indicate the means of the offset of the test sentence and the onset of the question. A baseline was calculated per trial. As the baseline varies per trial, it cannot be easily identified in the graph. See text for detailed information on the alignment and baselining of the data.
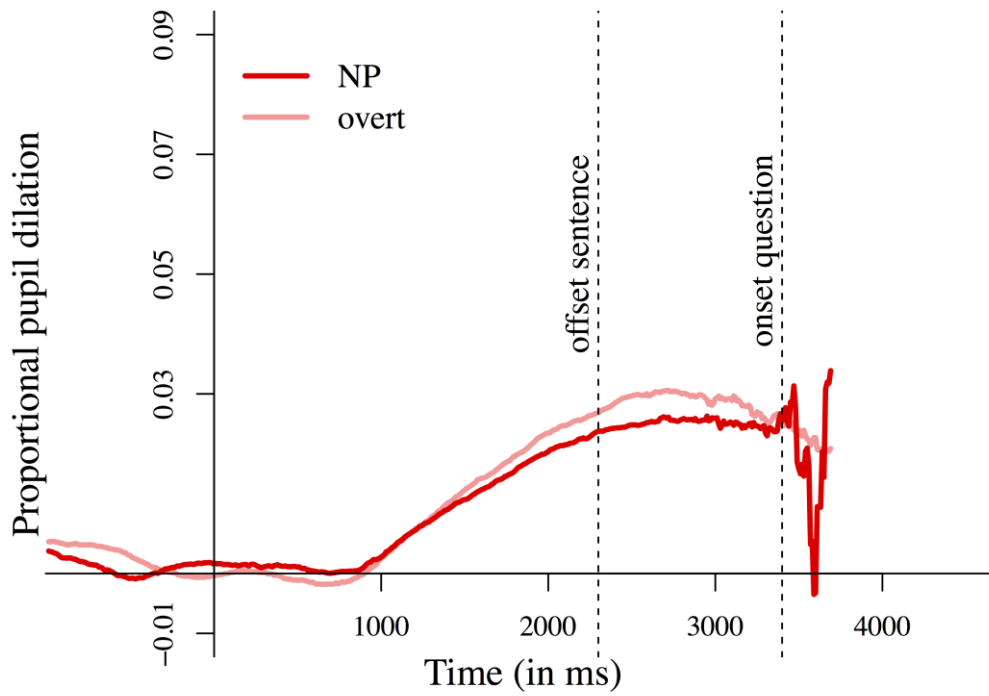
Figure 7. The mean proportional pupil dilation over time for the two anaphoric subjects in Dutch (NP: full NP subject, overt: overt pronoun subject). Time is aligned to the onset of the word following the subject. Vertical dotted lines indicate the means of the offset of the test sentence and the onset of the question.

## Appendix A: Results and analyses of the Italian data

Figure A1. The percentage of responses indicating a topic continuation for the three anaphoric subjects in Italian (NP: full NP subject, overt: overt pronoun subject, null: null subject pronoun) and the two anaphoric objects (reflexive, pronoun). Error bars are derived from logistic analysis.

Predictors	Estimate (β)	SE	z	p
Intercept	6.107	1.024	5.965	< 0.001
SubjectForm:overt	-6.574	1.028	-6.394	< 0.001
SubjectForm:null	-4.237	1.021	-4.150	< 0.001
ObjectForm:pro	-1.253	1.138	-1.101	0.271
SubjectForm:overt * ObjectForm:pro	1.034	1.154	0.896	0.370
SubjectForm:null * ObjectForm:pro	2.054	1.142	1.799	0.072

Table A1. An overview of the estimates (β), the *z*-scores, and the *p*-values of the fixed factors entered in the generalized linear mixed-effect model for the analysis of the responses in Italian (overt: overt pronoun subject, null: null subject, pro: pronoun object), with a baseline of a full NP subject and a reflexive object.

Predictors	Estimate (β)	SE	z	p
Intercept	-0.467	0.186	-2.520	0.012
SubjectForm:NP	6.574	1.028	6.396	< 0.001
SubjectForm:null	2.337	0.240	9.823	< 0.001
ObjectForm:pro	-0.219	0.246	-0.892	0.372
SubjectForm:NP * ObjectForm:pro	-1.034	1.154	0.896	0.372
SubjectForm:null * ObjectForm:pro	1.012	0.375	2.718	0.007

Table A2. An overview of the estimates (β), the *z*-scores, and the *p*-values of the fixed factors entered in the generalized linear mixed-effect model for the analysis of the responses in Italian (NP: full NP subject, null: null subject, pro: pronoun object), with a baseline of an overt pronoun subject and a reflexive object.
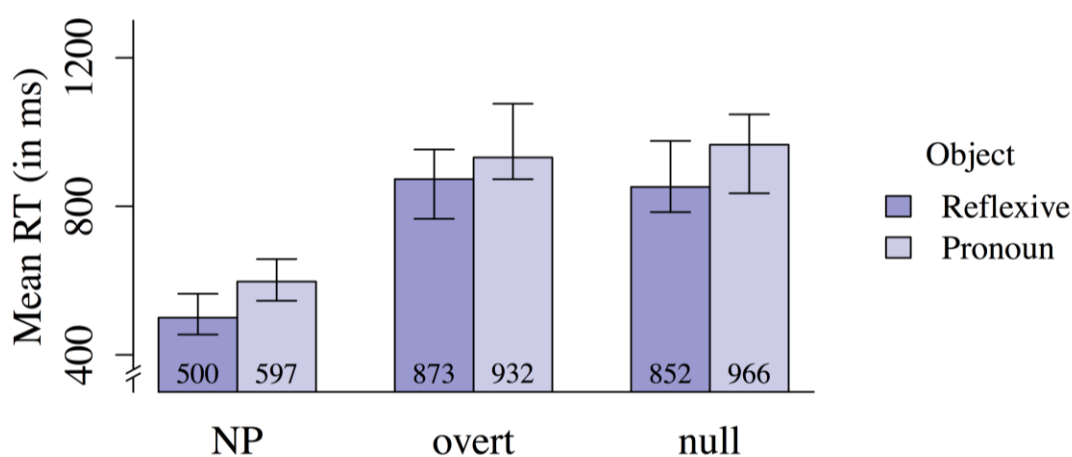

Figure A2. The mean response times for the interpretation questions regarding the three anaphoric subjects in Italian (NP: full NP subject, overt: overt pronoun subject, null: null subject) and the two anaphoric objects (reflexive, pronoun). Error bars are based on a 95% two-sided nonparametric confidence interval estimated through bootstrapping.

| Predictors	Estimate (*β*)	SE	df	*t*	*p*
Intercept	5.886	0.070	74	83.825	< 0.001
SubjectForm:overt	0.473	0.089	85	5.284	< 0.001
SubjectForm:null	0.487	0.093	77	5.207	< 0.001
ObjectForm:pro	0.172	0.076	1614	2.264	0.024
SubjectForm:overt * ObjectForm:pro	-0.121	0.107	1609	-1.128	0.259
SubjectForm:null * ObjectForm:pro	-0.078	0.107	1611	-0.732	0.465

Table A3. An overview of the estimates (*β*), and the *t*- and *p*-values of the fixed factors entered in the linear mixed-effect model for the analysis of the response times in Italian (overt: overt pronoun subject, null: null subject, pro: pronoun object), with a baseline of a subject NP and an object reflexive.

| Predictors	Estimate (*β*)	SE	df	*t*	*p*
Intercept	6.360	0.090	54	70.589	< 0.001
SubjectForm:overt	-0.473	0.089	85	-5.284	< 0.001
SubjectForm:null	0.014	0.081	116	0.176	0.860
ObjectForm:pro	0.051	0.076	1603	0.677	0.498
SubjectForm:overt * ObjectForm:pro	0.121	0.107	1609	1.128	0.259
SubjectForm:null * ObjectForm:pro	0.042	0.107	1603	0.396	0.692

Table A4. An overview of the estimates (*β*), and the *t*- and *p*-values of the fixed factors entered in the linear mixed-effect model for the analysis of the response times in Italian (overt: overt pronoun subject, null: null subject, pro: pronoun object), with a baseline of an overt subject pronoun and an object reflexive.

Pupil dilation data for each condition
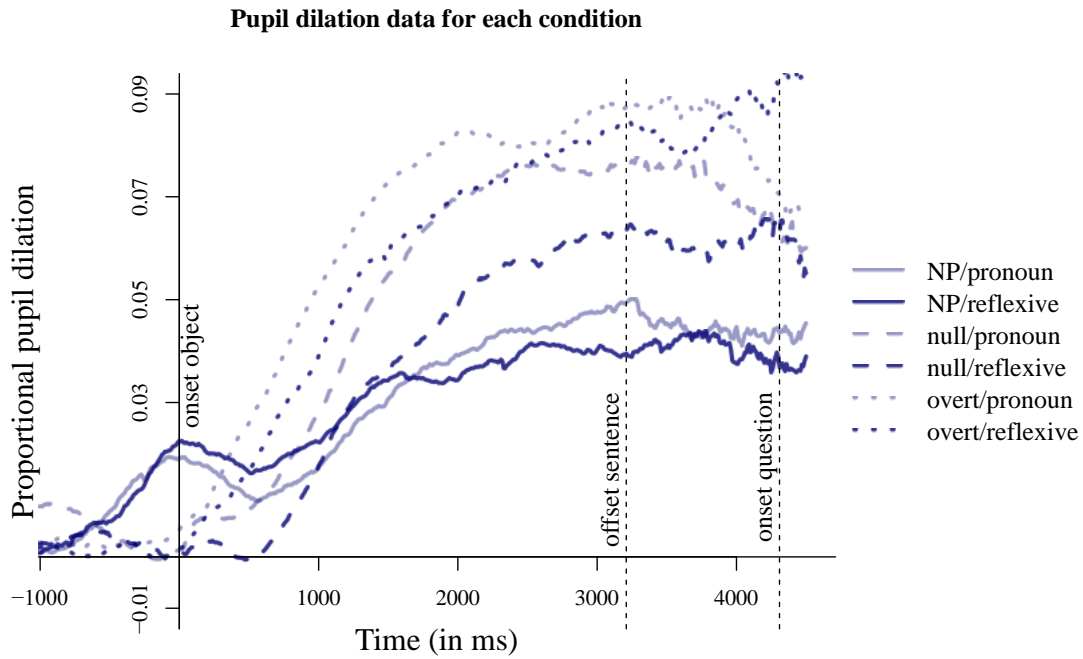

Figure A3. The mean proportional pupil dilation over time for each subject/object combination in Italian NP: full NP subject, overt: overt pronoun subject, null: null subject, pronoun: pronoun object, reflexive: reflexive object). Vertical dotted lines indicate the means of the offset of the test sentence and the onset of the question.

Pupil dilation data for the object forms
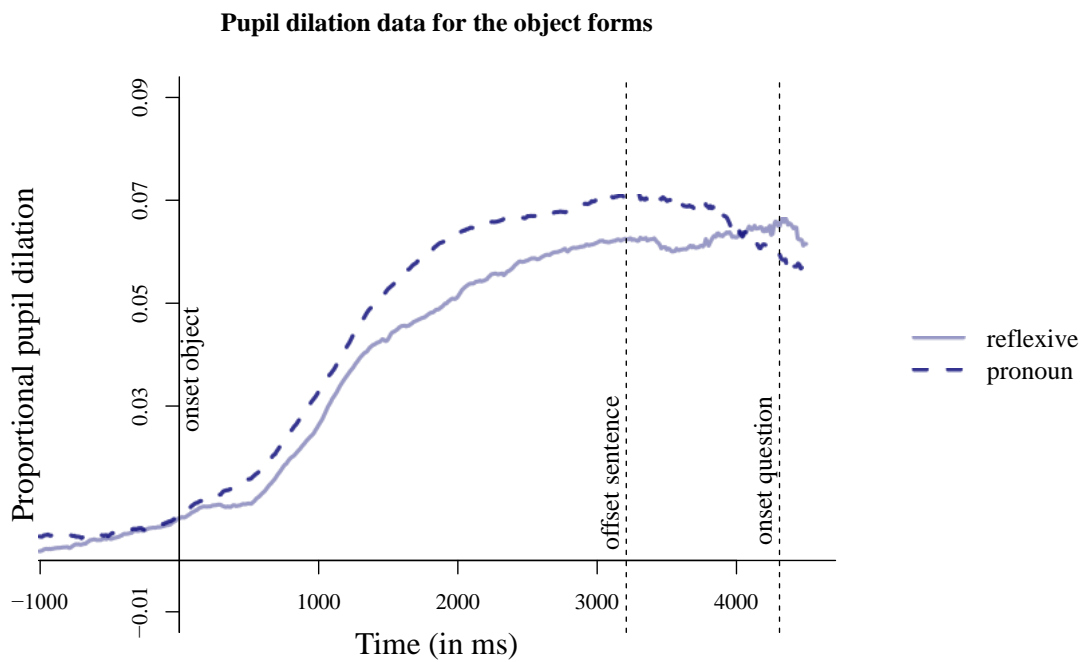
Figure A4. The mean proportional pupil dilation over time for the two object forms (reflexive and pronoun) in Italian. Vertical dotted lines indicate the means of the offset of the test sentence and the onset of the question.
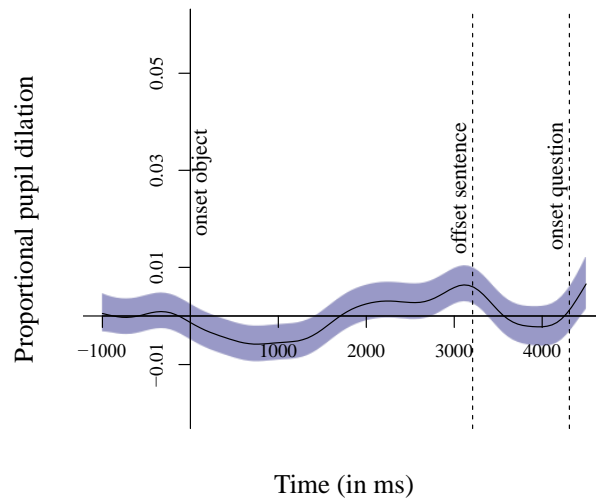
Figure A5. The model fits for the effect of a pronoun object compared to a reflexive object in sentences with a full NP subject (also see Table A5).
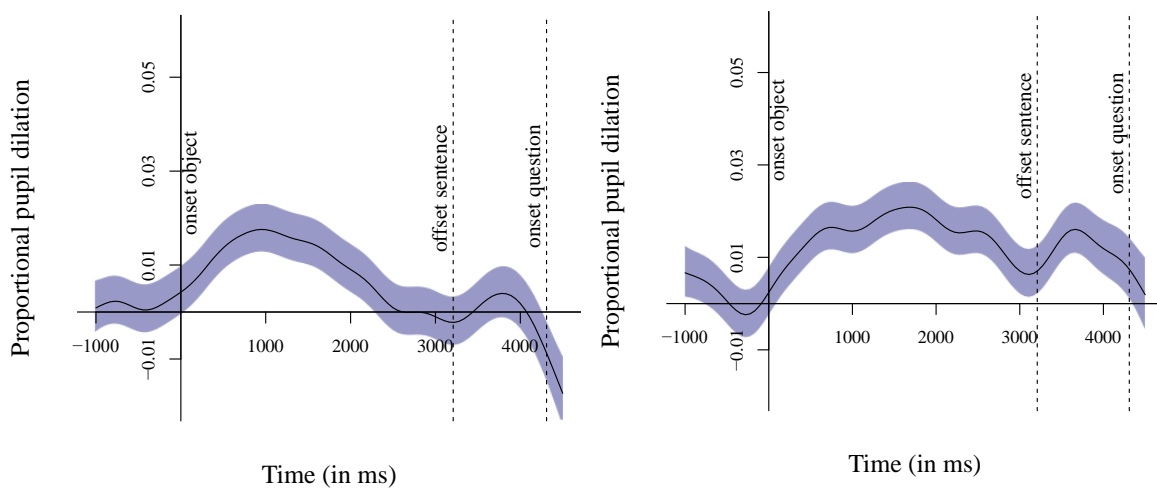
Figures A6 and A7. The model fits for the interaction between an overt pronoun subject and a pronoun object (left graph) and between a null subject and a pronoun object (right graph), as compared to the baseline of a full NP subject and a reflexive object (also see Table A5).

Parametric coefficients:

Predictor	Estimate (*β*)	SE	*t*	*p*
Intercept	0. 028	0.003	9.795	<2e-16

Approximate significance of smooth terms:

Smooth term	Edf	Ref.df	*F*	*p*
NP-reflexive baseline	16.77	18.156	13.393	< 2e-16
SubjectForm:null	17.12	18.889	22.004	< 2e-16
SubjectForm:overt	17.90	19.267	63.431	< 2e-16
ObjectForm:pro	17.21	18.804	4.252	1.76e-09
SubjectForm:null*ObjectForm:pro	16.71	18.646	9.206	< 2e-16
SubjectForm:overt*ObjectForm:pro	15.11	17.577	6.746	< 2e-16
s(Participant,Time)	38.79	39.000	219.108	< 2e-16
s(Item,Time)	92.77	95.000	50.799	< 2e-16
s(Trial)	8.87	8.992	53.866	< 2e-16
ti(Time,Trial)	64.95	73.658	9.955	< 2e-16

Table A5. Results of the GAMs analysis of the pupil dilation data, with a baseline of full NP subjects and reflexive objects (overt: overt pronoun subject, null: null subject, pro: pronoun object).

Parametric coefficients:

Predictor	Estimate (*β*)	SE	*t*	*p*
Intercept	0. 032	0.003	11.12	<2e-16

Approximate significance of smooth terms:

Smooth term	Edf	Ref.df	F	p
Null-reflexive baseline	17.21	18.349	31.994	< 2e-16
SubjectForm:NP	17.48	19.114	22.877	< 2e-16
SubjectForm:overt	16.92	18.771	25.877	< 2e-16
ObjectForm:pro	13.53	16.100	12.852	< 2e-16
SubjectForm:NP*ObjectForm:pro	17.00	18.819	9.274	< 2e-16
SubjectForm:overt*ObjectForm:pro	13.82	16.457	4.466	3.89e-09
s(Participant,Time)	38.79	39.000	219.124	< 2e-16
s(Item,Time)	92.77	95.000	50.811	< 2e-16
s(Trial)	8.87	8.992	53.867	< 2e-16
ti(Time,Trial)	64.95	73.656	9.963	< 2e-16

Table A6. Results of the GAMs analysis of the pupil dilation data, with a baseline of null subjects and reflexive objects (NP: full NP subject, overt: overt pronoun subject, pro: pronoun object).

Appendix B: Analyses of the differences between the Italian and the Dutch data

Predictors	Estimate (β)	SE	z	p
Intercept	4.949	0.441	11.209	< 0.001
Language:Dutch	-0.167	0.542	-0.308	0.758
SubjectForm:overt (Dutch)	-1.757	0.337	-5.216	< 0.001
SubjectForm:overt (Italian)	-5.399	0.431	-12.544	< 0.001
SubjectForm:null (Italian)	-2.844	0.432	-6.591	< 0.001
ObjectForm:pro	-0.135	0.127	-1.064	0.287

Table B1. An overview of the estimates (β), the z-scores, and the p-values of the fixed factors entered in the generalized linear mixed-effect model for the analysis of the responses in Italian and Dutch, with a baseline of Italian full NP subjects and reflexive objects (overt: overt pronoun subject, null: null subject, pro: pronoun object). The meaningful comparison in this model is the comparison of the baseline of Italian NP subjects to Dutch NP subjects, which is represented in the Language factor.

Predictors	Estimate (β)	SE	z	p
Intercept	2.163	0.193	11.229	< 0.001
Language:Dutch	0.787	0.245	3.207	0.001
SubjectForm:overt (Italian)	-2.614	0.165	-15.843	< 0.001
SubjectForm:NP (Italian)	2.272	0.263	8.626	< 0.001
ObjectForm:pro	-0.135	0.127	-1.068	0.286

Table B2. An overview of the estimates (β), the z-scores, and the p-values of the fixed factors entered in the generalized linear mixed-effect model for the analysis of the responses in Italian and Dutch, with a baseline of Italian null subjects and reflexive objects (NP: full NP subject, overt: overt pronoun subject, pro: pronoun object). The meaningful comparison in this model is the comparison of the baseline of Italian null subjects to Dutch overt pronoun subjects, which is represented in the Language factor.

Predictors	Estimate (β)	SE	df	t	p
Intercept	5.921	0.071	144	83.931	< 0.001
Language:Dutch	0.030	0.093	110	0.318	0.751
SubjectForm:overt (Dutch)	0.173	0.044	3214	3.915	< 0.001
SubjectForm:overt (Italian)	0.393	0.053	3220	7.403	< 0.001
SubjectForm:null (Italian)	0.437	0.053	3218	8.225	< 0.001
ObjectForm:pro	0.112	0.031	3228	3.615	< 0.001

Table B3. An overview of the estimates (β), and the t- and p-values of the fixed factors entered in the linear mixed-effect model for the analysis of the response times in Italian and Dutch, with a baseline of Italian full NP subjects and reflexive objects (overt: overt pronoun subject, null: null subject, pro: pronoun object). The meaningful comparison in this model is the comparison of the baseline of Italian NP subjects to Dutch NP subjects, which is represented in the Language factor.

Predictors	Estimate (β)	SE	df	t	p
Intercept	6.280	0.067	120	93.275	< 0.001
Language:Dutch	-0.103	0.086	82	-1.192	0.237
SubjectForm:overt (Italian)	0.034	0.049	3220	0.692	0.489
SubjectForm:NP (Italian)	-0.281	0.034	3217	-8.250	< 0.001
ObjectForm:pro	0.112	0.031	3229	3.615	< 0.001

Table B4. An overview of the estimates (β), and the t- and p-values of the fixed factors entered in the linear mixed-effect model for the analysis of the response times in Italian and Dutch, with a baseline of Italian null pronoun subjects and reflexive objects (NP: full NP subject, overt: overt pronoun subject, pro: pronoun object). The meaningful comparison in this model is the comparison of the baseline of Italian null pronoun subjects to Dutch overt pronoun subjects, which is represented in the Language factor.

Predictors	Estimate (β)	SE	df	t	p
Intercept	6.249	0.067	121	92.758	< 0.001
Language:Dutch	-0.081	0.086	82	-0.936	0.352
SubjectForm:null (Italian)	0.109	0.049	3220	2.218	0.027
SubjectForm:NP (Italian)	-0.263	0.034	3218	-7.733	< 0.001
ObjectForm:pro	0.112	0.031	3229	3.619	< 0.001

Table B5. An overview of the estimates (β), and the t- and p-values of the fixed factors entered in the linear mixed-effect model for the analysis of the response times in Italian and Dutch, with a baseline of Italian overt pronoun subjects and reflexive objects (NP: full NP subject,
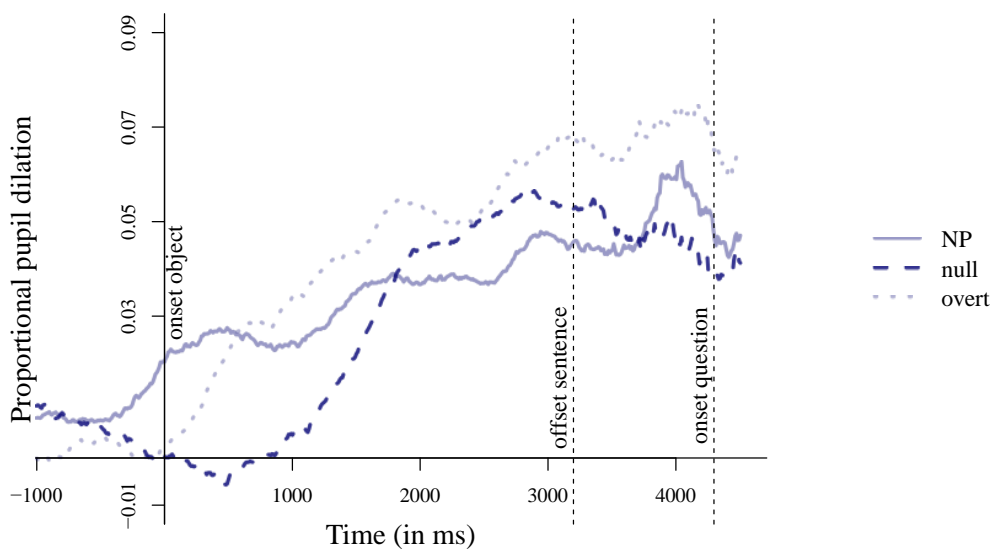
null: null subject, pro: pronoun object). The meaningful comparison in this model is the comparison of the baseline of Italian overt pronoun subjects to Dutch overt pronoun subjects, which is represented in the Language factor.

| Predictors	Estimate (β)	SE	t	p
Intercept	0.441	0.071	6.241	< 0.001
Language:Dutch	-0.414	0.112	-3.705	< 0.001

Table B6. An overview of the estimates (β), and the t- and p-values of the fixed factors entered in the linear mixed-effect model for the analysis of the slopes of the Italian and Dutch pupil dilation between 0 and 1000 ms after onset of the object, with a baseline of Italian. The meaningful comparison in this model is the comparison of the baseline of Italian to Dutch, which is represented in the Language factor.

Appendix C: Follow-up study with Italian and Dutch participants tested in the same lab

Pupil dilation is a measure that is very sensitive to internal as well as external influences. To check for influence of the specific location in which the participants were tested, we conducted a pilot experiment with the same setup as the described experiments, testing Italian participants (n=10) and Dutch participants (n=11) on their mother tongue in the same lab in the Netherlands. The results are plotted in Figure C1 and C2. Statistical analyses were not performed due to the low number of participants (and thus the low power) of the experiment.

Figure C1. The mean proportional pupil dilation over time for the three anaphoric subjects in Italian (NP: full NP subject, overt: overt pronoun subject, null: null subject) in the follow-up study. Time is aligned to the onset of the word following the subject. Vertical dotted lines indicate the means of the offset of the test sentence and the onset of the question.
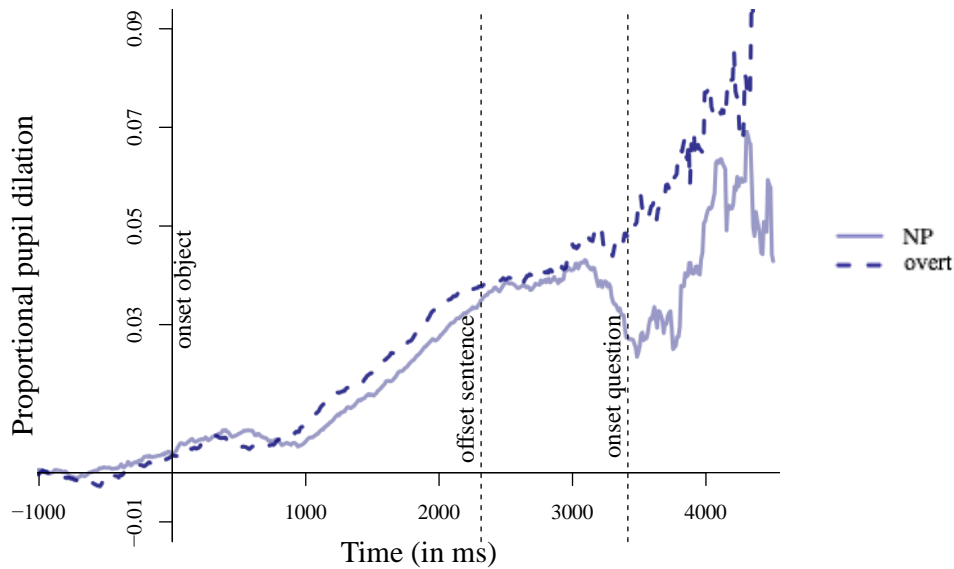
Figure C2. The mean proportional pupil dilation over time for the two anaphoric subjects in Dutch (NP: full NP subject, overt: overt pronoun subject) in the follow-up study. Time is aligned to the onset of the word following the subject. Vertical dotted lines indicate the means of the offset of the test sentence and the onset of the question.

Due to the low number of participants, the distinction between the conditions in the plots are less clear and the plots are less smooth than they could be. Nevertheless, Figure C2 shows pupil dilation effects similar to the ones reported in the paper, but, importantly, the amplitudes of the relative pupil dilation are much more similar between Italian and Dutch in this follow-up study than in the main experiment. In order to limit foreign language exposure contamination, we chose to test Italian participants in their home country for the main experiment.