Discourse Processes

Routledge
Taylor & Francis Group

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/hdsp20

Effects of event participant preview and patient animacy in sentence production: a cross-linguistic comparison between English and Russian

Johannes Gerwien, Judith Schlenter, Martina Penke & Agnieszka E. Konopka

To cite this article: Johannes Gerwien, Judith Schlenter, Martina Penke & Agnieszka E. Konopka (16 Dec 2024): Effects of event participant preview and patient animacy in sentence production: a cross-linguistic comparison between English and Russian, Discourse Processes, DOI: 10.1080/0163853X.2024.2433896

To link to this article: https://doi.org/10.1080/0163853X.2024.2433896

Published online: 16 Dec 2024.

DISCOURSE PROCESSES
https://doi.org/10.1080/0163853X.2024.2433896

Effects of event participant preview and patient animacy in sentence production: a cross-linguistic comparison between English and Russian

Johannes Gerwien[a], Judith Schlenter[b,c], Martina Penke[b], and Agnieszka E. Konopka[d]

[a]Institute of German as a Foreign Language Philology, Heidelberg University, Heidelberg, Germany; [b]Department of Special Education and Rehabilitation, University of Cologne, Cologne, Germany; [c]Department of Language and Culture, UiT the Arctic University of Norway, Tromsø, Norway; [d]School of Psychology, University of Aberdeen, Abeerden, UK

ABSTRACT

Theories of sentence planning (Linear vs. Hierarchical Incrementality) differ in their assumptions about how much conceptual information speakers require to initiate linguistic encoding. This picture description study tested whether speakers' selection of starting points for their sentences is influenced by the availability of information about a referent (perceptual accessibility) and by referent animacy (conceptual accessibility) in English and Russian, two languages differing in syntactic flexibility, case marking and other typological features. Target pictures showed transitive events with animate agents and animate/inanimate patients. One of the referents was previewed for 300 ms before presentation of the full picture. This preview manipulation was intended to enable earlier conceptual and lexical encoding of the first referent relative to the second referent. The frequency of agent-first structures and speech onset times (SOT) were compared between conditions as well as across the course of the experiment. The results showed that the likelihood of speakers producing agent-first responses (1) dropped when patient referents are previewed, (2) dropped when patient referents are animate, and (3) changed over the course of the experiment in English and Russian in different ways. Analyses of speech onsets of agent-first sentences also showed increases in processing costs in conditions where animate patients were previewed, consistent with a revision process taking place after early encoding of a referent that was not ultimately produced in sentence-initial position. Taken together, the findings suggest that both speaker groups engaged in linearly incremental encoding (i.e. reliance on minimal conceptual information to begin planning) to some extent, and that reliance on this planning strategy can change over time. The results are also discussed in the context of the Production-Distribution-Comprehension account.

Introduction

Production of a multiword utterance requires outputting words one by one, and each utterance must have a *starting point*. For example, to describe an event in which a man smashes a window, speakers may choose to convey the gist of this event in a number of ways. If they produce a transitive sentence, they may begin with "The man . . . " (e.g., "The man is smashing the window") or with "The window . . . " (e.g., "The window is being smashed by the man"). Selecting a starting point has implications for the linear order of subsequent content words and for the syntactic structure of the

CONTACT Agnieszka E. Konopka agnieszka.konopka@abdn.ac.uk

sentence. In languages with fixed word order like English, beginning a transitive sentence with an agent or a patient commits the speaker to use a particular syntactic structure (active or passive syntax) and thus implies selection of a sentence subject. In languages with free word order, like Russian, selection of starting points and selection of syntactic structures can be dissociated to some extent: active sentences may, for example, start with a direct object and include the sentence subject in sentence-final position.

How do speakers decide where to start? Understanding the processes involved in the selection of starting points has wide-ranging implications for theories of sentence planning – in particular for the relationship between conceptual and linguistic encoding. Production begins with the generation of a preverbal message and is followed by selection of message-appropriate lexical items and syntactic structures to express this message linguistically (Levelt, 1989). However, although it is clear that some conceptual planning must occur before linguistic encoding can be initiated, existing theories differ in their assumptions about the nature and breadth of the information planned at the outset of formulation at the conceptual level. Is one referent concept enough (e.g., *man* or *window*) to begin the linguistic encoding process, or do speakers need more information (e.g., the concepts *window, man*, and *smash*, with some information about the relationship between these concepts)? Here, we address the question of whether information availability influences production and the degree to which early conceptual and linguistic planning involve relational (i.e., structural) processing in two languages: English and Russian.

Linear and *Hierarchical Incrementality* are theoretical frameworks that describe the relationship between planning processes at the message level and at the sentence level. They make contrasting predictions about the size of the advance planning units at the conceptual level (one concept vs. multiple concepts encoded in a relational, message-level framework respectively) as well as about the ensuing lexical and syntactic encoding operations at the sentence level. These predictions are perhaps articulated most clearly in the debate about starting points.

According to *Linear Incrementality*, speakers may simply encode first what they think about or what they see first, or – in more general terms – what is most salient at a given moment in time (Bock et al., 2004). In other words, the temporal availability of conceptual information can guide the selection of starting points: the most salient concept at the message level becomes, by default, the starting point for the sentence. Compelling demonstrations in favor of linearly incremental planning come from studies using perceptual cueing paradigms (e.g., Tomlin, 1995, 1997). In Tomlin's well-known *Fish Film* paradigm, English speakers described scenes with two fish, one of which (the agent) was about to eat the other (the patient). An exogenous cue (an arrow) was used to direct speakers' attention to one of the two fish. When the arrow pointed to the agent fish, participants started their utterances with the agent (active sentences), but this bias shifted when the arrow pointed to the patient fish: in these cases, speakers almost exclusively started their utterances with the patient, resulting in the production of a passive. Importantly, the strong version of Linear Incrementality proposes that selecting a starting point does not need to entail any syntactic planning: speakers need a preverbal message consisting of as little as a single concept to launch lexical encoding and they can engage in conceptual and syntactic planning for the rest of the sentence *after* the starting point has been selected (Bock & Ferreira, 2014; Konopka & Brown-Schmidt, 2014).

An alternative to linear, availability-driven planning strategies is hierarchical planning. *Hierarchical Incrementality* proposes that speakers need a broader conceptual representation before initiating encoding and thus that the selection of a starting point is influenced by a number of factors, including the ease of relational encoding (e.g., encoding relationships between referents, generating structures), discourse continuity, and pragmatics, rather than by a simple linear assignment of the first-encoded concept to a sentence-initial position. Support for *Hierarchical Incrementality* comes from studies in which speakers are being eye-tracked while they produce verbal descriptions in response to scenes showing simple events. Findings show that speakers first engage in rapid gist extraction (<300 ms) during which they establish relations between individual characters, and, on the basis of that representation, they then direct their gaze to the referent that they will mention first. Such findings have been interpreted to show that an initial conceptual representation of the relationships between referents in a scene is generated first, and that speakers select a starting point based on this representation (Bock et al., 2004; Griffin & Bock, 2000). The selection of a starting point is thus not motivated by the immediate salience of individual referents, but is rather characterized as greater advance planning at the conceptual level (including event structure).

Importantly, gaining insight into the planning process – and thus pitting the predictions of *Linear* versus *Hierarchical Incrementality* against each other – requires overcoming at least one substantial methodological challenge. The most common approach to studying language production is to elicit responses from *visual* stimuli, such as pictures of objects or pictures of simple events. Such stimuli, however, provide speakers with all the information they need to produce a sentence at once. The timing of gist encoding in such stimuli is also debated (e.g., Hafri et al., 2013). Thus, experimental studies have used a range of manipulations to track what *type* of information in the visual stimuli influences speakers' linguistic choices, particularly at the outset of the planning process.

Referent animacy, cueing, and preview manipulations

Stimulus pictures of simple events normally show interactions between a range of different referents, and properties of these referents can be manipulated to assess their influence on the time-course of planning. Manipulations of referent animacy are motivated by the fact that the suitability of individual event characters (e.g., agents and patients) for starting points of short sentences varies with conceptual features that are tied to accessibility. Animate referents are inherently more salient and accessible than inanimate ones (e.g., Bock & Warren, 1985; Prat-Sala & Branigan, 2000; Van Nice & Dietrich, 2003), which makes them more likely starting points than inanimate referents (all else being equal). Animacy is also associated with agenthood (although both agents and patients in transitive events can, of course, be animate or inanimate; e.g., Bornkessel-Schlesewsky & Schlesewsky, 2009; Branigan et al., 2008). This link between animacy and agenthood has immediate implications for the selection of starting points (e.g., Tomlin, 1986). In many languages, animate agents typically precede inanimate patients in canonical active sentences. Correspondingly, speakers are more likely to place an animate than an inanimate patient in sentence-initial position, which results in the production of a noncanonical passive sentence. Such choices suggest that identification of a referent as being animate or inanimate involves some degree of weighting of its suitability for the sentence subject role, which is more consistent with Hierarchical Incrementality (where relational processing is at the core of early planning decisions) than Linear Incrementality (where relational processing is assumed not to influence early planning). However, the mechanisms implicated in these decisions (e.g., the degree to which animacy influences linear word order vs. grammatical function assignment) is still debated (Branigan et al., 2008), as is the degree to which animacy can trigger early relational processing relative to other variables influencing the salience of a referent.

The extent to which starting point choices can be manipulated has also been investigated using perceptual cueing techniques, in which speakers' initial visual attention is systematically drawn to different parts of a display. For example, the presentation of a target picture showing an agent acting on a patient may be preceded by a cue shown in the location of either event participant. Cues can be genuinely perceptual (e.g., dots) or referential (e.g., referent previews), and can vary in duration: subliminal (or implicit) cues are presented briefly enough to be imperceptible, while explicit cues are presented for several hundred milliseconds (ms) and are consciously perceived. The results of studies using such manipulations consistently showed that referent cueing increases the likelihood of this referent becoming the starting point of a sentence: e.g., English speakers produce more passive utterances after patient cues than after agent cues, irrespective of cue type (dot vs. referential cue: Myachykov et al., 2012) and cue duration (subliminal/implicit: e.g., Gleitman et al., 2007; Hwang & Kaiser, 2015; Myachykov et al., 2018; explicit cueing: e.g., Myachykov et al., 2012, 2018). Thus, manipulating referent salience or accessibility via nonlinguistic cues shows that simply orienting a speaker's attention and providing information about referents in different orders can reliably influence the selection of starting points and linear word order in simple sentences.

Comparisons of speech onset latencies across cueing conditions support these conclusions. For example, English sentences starting with an agent have shorter latencies after agent cues than patient cues (Myachykov & Tomlin, 2008). Moreover, latencies were found to be shorter after explicit referential than after simple perceptual cues, suggesting more facilitation after a conceptually informative referential cue (Myachykov et al., 2018). Similarly, explicit referential patient cueing results in shorter latencies for passives than simple perceptual patient cueing (Schlenter et al., 2022), and explicit referential patient cueing results in shorter latencies for passive than active utterances in German (Esaulova et al., 2020; Schlenter & Penke, 2022; for differential effects based on cue type, see also Dolscheid & Penke, 2023).).

Cross-linguistic differences in selection of starting points

The debate about the suitability of linearly incremental and hierarchically incremental planning strategies for sentence production is also informed by cross-linguistic comparisons. These comparisons have shown interesting differences in the degree to which speakers' linguistic choices can be manipulated by cueing and previewing techniques. For example, Myachykov and Tomlin (2008) used the *Fish film* paradigm to elicit event descriptions in Russian, and showed that, although patient cueing increased the production of patient-initial utterances, the cueing manipulation was less effective than in English: only 36% of the experimental items with a cued patient were described with patient-initial utterances (compared to virtually 100% in English). Furthermore, patients were selected as starting points more often than in English across cueing conditions, but not as sentence subjects. Similarly, when presented with an explicit (500 ms) cue in the location of the patient prior to display onset, Russian speakers preferred to produce an object-initial active sentence over a passive one (Pokhoday et al., 2019). Importantly, this was the case even though Russian is much more flexible in terms of word order compared to English (Wade, 2011). Specifically, Russian allows six different word orders for simple transitive sentences.[1]

One explanation for the cross-linguistic difference observed in cueing studies relates to case marking and structure frequencies. On the one hand, case marking typically goes hand in hand with more syntactic options in word order. On the other hand, languages with overt case marking have been shown to allow linearly incremental encoding to a much lesser extent than languages without overt case marking, like English (Egurtzegi et al., 2022; Hwang & Kaiser, 2015; Norcliffe et al., 2015; also see Momma et al., 2016, and Sauppe et al., 2013). This is because overt case-marking often requires early assignment of syntactic functions (subject, object) during planning, which relies on relational information specified at the conceptual level (i.e., thematic role assignment). Additionally, even if a language allowed the same thematic role to be mapped onto different syntactic functions (e.g., a patient as a subject or direct object) and to be realized in several sentence positions (e.g., sentence-initial and sentence-final positions), having more syntactic options in general does not necessarily imply more syntactic freedom under all circumstances. Russian speakers, for example, can choose from a large inventory of syntactic structures, but these structures are subject to specific discourse constraints and may be associated with a specific register or style. The passive voice in Russian, for example, is usually dispreferred outside appropriate contexts (e.g., pragmatically rich contexts where the information structure of the message strongly biases speakers to select passive constructions) and is generally rather infrequent in informal spoken language (e.g., Myachykov et al., 2011). Thus, in experiments where participants produce isolated sentences under time pressure, speakers of a language that in principle offers a variety of syntactic options often do not make use of this flexibility, but instead repeatedly use a subset of all possible structures.

At the same time, the availability of a wider range of syntactic options can influence the time-course of the planning and production process, even when it does not affect the production outcome. For example, Myachykov et al. (2013) found that speech onset latencies were longer in Russian than in

English. Similarly, Hwang and Kaiser (2014) found that Korean speakers compared to English showed a slow-down in sentence production when they could choose between alternative syntactic function assignments. These findings suggest that choosing between structural alternatives (choosing between syntactic frames or choosing between different function assignment options) may be costly.

The evidence reviewed here shows that the selection of starting points can be influenced by a number of factors, including the mere availability of information about a single referent (which is consistent with Linear Incrementality) on the one hand and conceptual properties of this referent as well as language-specific preferences (which is less consistent with Linear Incrementality and more consistent with Hierarchical Incrementality) on the other hand. Overall, such differences both within and across languages align better with the proposal that speakers can rely on linearly incremental and hierarchically incremental planning to different extents under different conditions, i.e., with the proposal that Linear and Hierarchical Incrementality represent two end points on a continuum of plausible planning strategies (e.g., Konopka & Meyer, 2014; Van de Velde et al., 2014), and that the choice to rely on either strategy can be driven, in the moment, by properties of the to-be-encoded message and by properties of the target language. An important direction for cross-linguistic research on message and sentence planning is therefore the need to clarify (a) whether properties of a language are generally more likely to support one type of planning strategy over another (i.e., to produce stable cross-linguistic differences) and (b) whether there are conditions under which speakers of different languages employ similar strategies (i.e., to produce stable cross-linguistic similarities, consistent with either Linear or Hierarchical Incrementality). However, substantial differences in methodologies across previous studies prevent the drawing of strong conclusions based on typological differences. To this end, we conducted a production study with speakers of five different languages using the same experimental design and procedure. This enables a direct investigation of cross-linguistic previewing effects, allowing us to shed light on the interaction of factors such as case marking, word order flexibility, and structure frequency, as well as gauging whether different languages might systematically favor planning strategies falling at different points on the continuum between Linear and Hierarchical Incrementality. In the present article, we focus on the comparison between English and Russian.

In line with previous studies, we interpret the current results in terms of a contrast between greater reliance on linearly incremental planning vs. greater reliance on hierarchically incremental planning. As language production is subject to general cognitive processing costs and constraints and speakers can select strategies that allow them to mitigate these costs, we then contextualize our findings by expanding the discussion with reference to the Production-Distribution-Comprehension account (MacDonald, 2013) We show how differences in planning strategies attributed to message-specific and language-specific differences are compatible with the (1) *Easy First*, (2) *Reduce Interference*, and (3) *Plan Reuse* principles outlined by MacDonald (2013).

Current study

Native speakers of English and Russian spontaneously described pictures depicting a range of cartoon-like scenes (e.g., Esaulova et al., 2019; Konopka et al., 2018). We employed a referent preview manipulation where on each trial, participants were given a 300-ms preview of one character or object prior to seeing the entire picture. On target trials, the pictures showed two-referent transitive actions with an animate agent acting on an animate or inanimate patient. Thus, speakers received a preview of either the agent or patient, and the patient could be either animate or inanimate. The preview manipulation was expected to allow for conceptual encoding of the previewed referent to begin ahead of the second referent. The preview time of 300 ms is rather short compared to previous studies, but it should be sufficient for the activation of a lexical concept and its corresponding lemma (Gleitman et al., 2007; Indefrey & Levelt, 2004). However, the final processing steps for the previewed referent (phonological/phonetic encoding and initiation of articulation) will unavoidably coincide with the initial processing of information that becomes available with full picture onset. Thus, two general scenarios can be expected in our study, based on the brief preview time manipulation: (1) the encoding of the previewed referent as a possible starting point proceeds when the entire scene becomes available (with all event structure information), or (2) the encoding of the previewed referent is temporarily interrupted when the full scene becomes available and thus when speakers can identify a more suitable starting point. The first scenario is predicted by Linear Incrementality, the second one by Hierarchical Incrementality (see later for more specific predictions). Based on judgments provided by native speakers, we initially expected to see the structures listed for English and Russian. However, our later results showed much less variability in the Russian data (see Table 1).[2]

(1) English
a) Agent-first, active AVP (Agent-Verb-Patient)
A bee is stinging a man/A bee stings a man
b) Patient-first, passive PVA (Patient-Verb-Agent)
A man is being stung by a bee

(2) Russian

a. Agent-first, active AVP
Пчела ужалила мужчину.
bee.NOMsting.PERF.PST man.ACC
"A bee stung a man."[3]

b. Patient-first, active PVA
Мужчину ужалила пчела.
man.ACC sting.PERF.PST bee.NOM
"A bee stung a man."

c. Agent-first, active APV
Пчела мужчину ужалила.
bee.NOMman.ACC sting.PERF.PST
"A bee stung a man."

d. Patient-first, active PAV
Мужчину пчела ужалила.
man.ACC bee.NOM sting.PERF.PST
"A bee stung a man."

e. Patient-first, passive PVA
Мужчина был ужален пчелой.
man.NOM was sting.PERF.PTCP.PASSbee.INS
"A man was stung by a bee."

f. Agent-first, passive AVP
Пчелой был ужален мужчина.
bee.INS was sting.PERF.PTCP.PASS man.NOM
"A man was stung by a bee."

Using the experimental setup described here, we tested the limits of linearly incremental planning in the two languages as follows. *First*, we assessed the extent to which the assignment of a referent to sentence-initial position is mediated by perceptual availability (i.e., referent preview) and conceptual accessibility (i.e., animacy), and whether the effects of these manipulations are different for referents with different event roles (i.e., agents vs. patients). Under Linear Incrementality, speakers should produce any accessible referent in sentence-initial position, whether this accessibility is perceptual or conceptual in nature. In other words,

Table 1. Distribution of syntactic structures in English and Russian in the agent-preview and patient-preview conditions (truncated passives in parentheses).

		Agent Preview		Patient Preview	
Language	Structure	Animate Patient	Inanimate Patient	Animate Patient	Inanimate Patient
English	active (SVO)	276	265	223	242
English	passive (SVO)	10	0	57 (3)	11 (1)
English	relative clause	0	0	1	0
English	Total	286	265	280	253
Russian	active (SVO)	258	299	240	276
Russian	active (OVS)	9	1	15	6
Russian	passive	0	0	4 (3)	1
Russian	passive (3rd Pl)	2	0	4	5
Russian	relative clause	0	0	0	2
Russian	Total	269	300	263	290

previewed referents should be mentioned first, regardless of whether they are animate or inanimate and regardless of whether they then turn out to be agents or patients. On a weaker version of Linear Incrementality where conceptual accessibility may play a more important role than perceptual accessibility, previewing an *animate* referent may trigger provisional grammatical/thematic role assignment and facilitate production of sentences beginning with this referent. Thus, the effects of referent animacy may be enhanced by referent preview, possibly to a different extent for agents and patients. Speakers generally prefer to produce animate-first than inanimate-first sentences and agent-first than patient-first sentences. So, previewing an animate character that then turns out to be an agent is unlikely to increase the production of agent-first sentences any further (Bock & Levelt, 1994; Christianson & Ferreira, 2005). In contrast, previewing an animate character that then turns out to be a patient may have a stronger effect on the selection of starting points, resulting in more patient-first sentences (e.g., Esaulova et al., 2021). *Second*, we tested the generalizability of findings across two languages differing in case marking, in the number of plausible syntactic options (and thus syntactic flexibility), and the degree to which context constrains the use of noncanonical structures like the passive. We expected to replicate previous findings regarding the placement of animate referents in sentence-initial position in English and Russian, but with differences in sensitivity to the referent preview manipulation across languages – specifically, weaker effects of patient preview on selection of starting points in Russian than in English (MacDonald, 2013).

For a finer-grained analysis of planning processes, we also compared speech onset times (SOTs) across conditions for agent-first sentences (i.e., the preferred sentence type with the largest number of responses). SOTs reflect the degree of difficulty in initiating sentences with specific referents in sentence-initial position. We expected shorter onset latencies after agent previews than after patient previews in both languages, as speakers can proceed with the production of a canonical agent-verb-patient (AVP) sentence after an agent preview (e.g., Myachykov et al., 2018), yet the production of an AVP sentence after a patient preview requires viewing the full scene (after the 300-ms preview phase) or may require some replanning. The predictions regarding animacy are less clear-cut, as previous studies in English and Russian did not systematically test the influence of patient animacy on sentence planning within a cueing or preview design. We may assume that, in cases where an active SVO sentence (i.e., AVP) is produced on a patient preview trial, SOTs are longer in sentences with an animate patient than an inanimate patient. This is because speakers may initially assume that animate referents are likely to be agents, and then, upon seeing the full scene, they need to interrupt and revise the current plan in order to produce a sentence with the second referent in sentence-initial position instead. Finally, based on the predictions of the syntactic flexibility account, we predicted longer SOTs in Russian than in English, because Russian speakers have more structural alternatives. So far, studies carried out with speakers of English have included simple perceptual cues (e.g., Gleitman et al., 2007; Hwang & Kaiser, 2015; Myachykov et al., 2011, 2018) and referent preview (Myachykov et al., 2018, 2012), yet the findings for Russian are limited to concurrent perceptual and referential cueing in the *Fish Film* paradigm (Myachykov & Tomlin, 2008) and to perceptual cueing prior to picture onset (Pokhoday et al., 2019), as no study investigated the effect of referent preview alone. Previous studies in English and Russian also did not systematically test how referent animacy (patient animacy in particular) influences production.

Method

Participants. A total of 40 English and 40 Russian speakers participated in the study. English speakers (age: $M = 23.8$, $SD = 5.3$; 11 males, 29 females) were students at the University of Aberdeen and received course credit for participation. Russian speakers (Age: $M = 24.8$, $SD = 6.7$; 8 males, 32 females) resided in Russia or Ukraine at the time of the data collection. They were recruited via word of mouth and participated voluntarily. A pre-experimental questionnaire asked participants to indicate whether they had acquired English or Russian respectively as (one of their) first languages, i.e., between the ages of 0 and 6 (two English and six Russian participants who did not meet these criteria were excluded).

Materials and design

The materials consisted of 40 target pictures and 40 fillers. The target pictures showed transitive events: half showed an animate agent acting on an animate patient and half showed an animate agent acting on an inanimate patient. Most items had been used in previous studies (in different languages) and thus were expected to reliably elicit transitive descriptions. Filler pictures were intended to elicit descriptions that were conceptually and syntactically different from the targets: they showed spatial configurations of objects (e.g., a balloon over a fence; $N = 17$), intransitive events where a person or object was involved in some activity or state (e.g., a man jumping rope, a candle burning; $N = 21$), and transfer-of-possession events (e.g., a woman throwing a frisbee to a girl; $N = 2$) and were also selected from previous studies, web searches, and from the MultiPic database (Duñabeitia et al., 2018). All pictures were presented in grayscale and varied in style and level of detail.

In addition to the between-item variable Patient Animacy, we factorially manipulated the preview of one of the two referents (agent or patient) within items. The placement of the agent relative to the patient (left vs. right-hand side of the screen) was counterbalanced, so that participants could not assign a thematic role to the previewed referent based on its screen position. All animate referents faced each other (see Figure 1). The 40 target items were distributed equally across two presentation lists, so that each participant encountered a scene only once in one of the two preview conditions.

Experimental procedure

The experiment was built with a PennController 1.9 and run on PCIbex Farm (Zehr & Schwarz, 2018). Participants received a link to the experiment and took part from home. They were assigned to one of the two stimuli lists (counterbalancing preview condition) at random. At the start of the experiment, they gave informed consent and granted access to their microphone. Next, they answered demographic questions (age, gender, handedness, first language(s)) and received instructions. They were asked to describe pictures of simple events with one sentence, but they were informed that they would see only a part of the full picture first for a brief amount of time, followed by the presentation of the full picture. They were instructed to begin speaking as soon as possible and to avoid using fillers like *uhm, ehm*, etc. As the experiment was web-based, no further instructions were provided during the session.

Each trial started with the presentation of a central fixation cross for 1 second. Next, participants saw one of the referents in the upcoming event for 300 ms. This was followed by presentation of the full picture for 2 seconds and then a blank screen for a further 2 seconds, to allow participants to complete their responses. After the blank screen, participants clicked on a button to continue. The

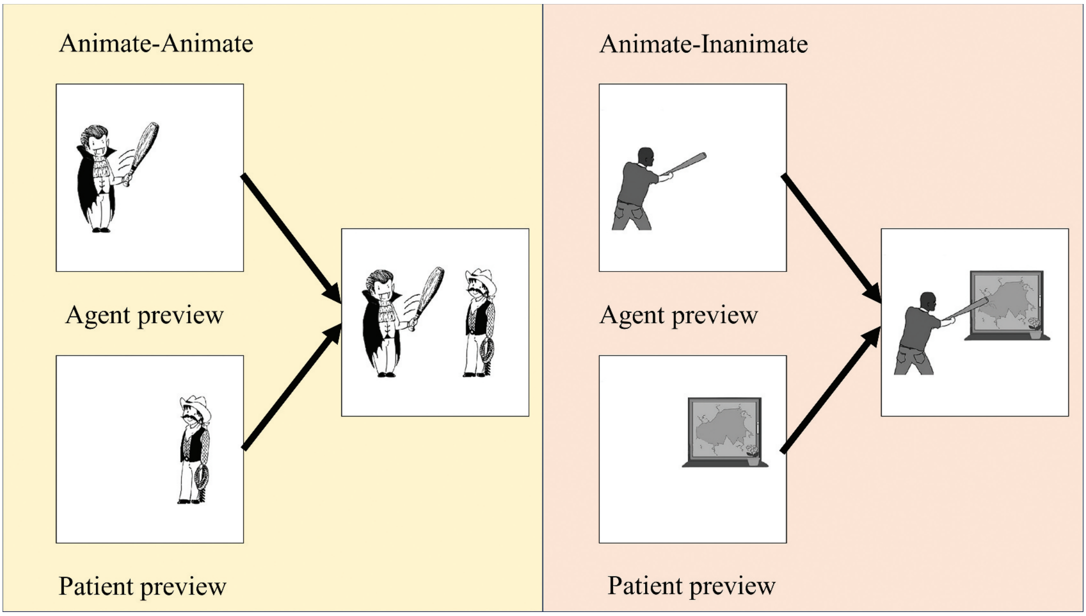

Figure 1. Example target picture showing an animate agent and animate patient (left) and an animate agent and inanimate patient (right). Scenes were preceded by a 300 msec preview of the agent or patient.

audio recording started with the disappearance of the fixation cross and stopped with the mouse click. Trial order was randomized on a by-participant basis. Altogether, the experiment lasted about 15 minutes.

Scoring and analysis

Participants' responses were transcribed including all pauses, hesitations, and other filler words. Speech onset times were manually marked using PRAAT (Boersma, 2001). Responses from target trials were coded with respect to utterance content and structure in a multistep process. In the first step, all responses that did not include a transitive verb referring to the action depicted in the picture were discarded.[4] This included responses that mentioned only one of the two referents ("A pirate"), mentioned both referents in a complex noun phrase with no verb ("A pirate and a fisherman"), or described the referents' position relative to one another ("A pirate [is/is standing] next to a fisherman"). In the remaining dataset, there were three types of valid responses: active sentences with a direct object ("A cat is catching a mouse") or with a prepositional object ("A cat is playing with a mouse"), and passive sentences with or without an agent[5] ("A mouse is being caught by a cat," "A mouse is being caught"). We also included "reduced" (or "telegraphic") utterances that omitted auxiliaries or determiners ("(A) cat (that is) catching a mouse"). Only responses with speech onset times lower than 4000 ms were included.

In a second step, we assessed the number of valid responses per participant. We required participants to provide at least 20 valid trials (i.e., scorable responses on at least 50% of target trials) according to the aforementioned criteria. Our rationale was that if participants produced fewer valid responses, they had likely not fully understood the task, were not appropriately engaged, or had internet connectivity problems. We excluded 6 participants in total, leaving data from 36 English speakers (mean number of valid responses = 32.8, *SD* = 4.91) and 38 Russian speakers (mean number of valid responses = 31.8, *SD* = 3.95). Because our goal was to compare responses across languages for the same stimuli, we then checked the number of valid responses per picture in each language to identify discrepancies. We assessed whether all stimuli elicited valid responses from at least 50% of the remaining participants in each language. Five pictures did not fulfill this criterion and were excluded (3 from the inanimate-patient and 2 from the animate-patient category).

Applying all exclusion criteria left 2204 trials (English: 1084, Russian: 1120) for further analysis. Most responses (99.2%) included two referent names (agent and patient). Truncated passives ("A football is being kicked") made up only 0.8% of the data. Table 1 shows a summary of the syntactic structures produced across all valid trials, and Figure 2 shows by-participant proportions of agent-first responses[6] and speech onset times.

Results

Starting points – linear order of agents and patients

We used the *brms* R package (Bürkner, 2017) to run Bayesian generalized mixed models in R (R Core Team, 2022, version 4.2.1) assuming a Bernoulli distribution with a logit link function to estimate the distributional properties of agent-first responses across conditions and language groups, as well as all interactions. Regression-style coding was applied for the two levels of all predictors (Preview Condition: – 0.5 = patient preview, 0.5 = agent preview; Patient Animacy: – 0.5 = animate, 0.5 = inanimate; Language: – 0.5 = English, 0.5 = Russian). The model included random interacting by-participant slopes for Preview Condition and Patient Animacy and by-item interacting slopes for Preview Condition and Language.[7] Parameters for the Bayesian models were fitted with four Markov Chain Monte Carlo (MCMC) chains. Each chain contained 2000 burn-in samples and 12000 additional samples, resulting in 10000 posterior samples per chain that were combined to one posterior sample consisting of 40000 samples for each model parameter. Model convergence was evaluated based on the Gelman–Rubin convergence statistic R-hat (Gelman & Rubin, 1992). Model fit was evaluated by comparing samples from the posterior predictive distribution to the observed responses.

Given that we have prior knowledge about cueing and preview effects from earlier research (i.e., we know that these manipulations are generally effective but can differ in magnitude across languages), we assumed that the probability of producing an agent-first response would not be 0.5 (this would be expected if preview effects were observed on all agent-preview and patient-preview trials). Considering potential modulation by patient animacy as well as the relatively brief preview duration, assuming a 0.9 probability of agent-first responses seemed like a reasonable choice. These considerations are reflected in the prior on the overall intercept (normal(2, 0.1)). Regression weights had normally distributed priors (normal (0, 1)), standard deviations of random effects had student t distributed priors, $\sigma \sim$ student_t(3, 0, 2.5), and the covariance matrices had LKJ Cholesky hyperpriors with $\eta = 1$. In addition to the model just described, we set up a second model with the same specifications but included trial

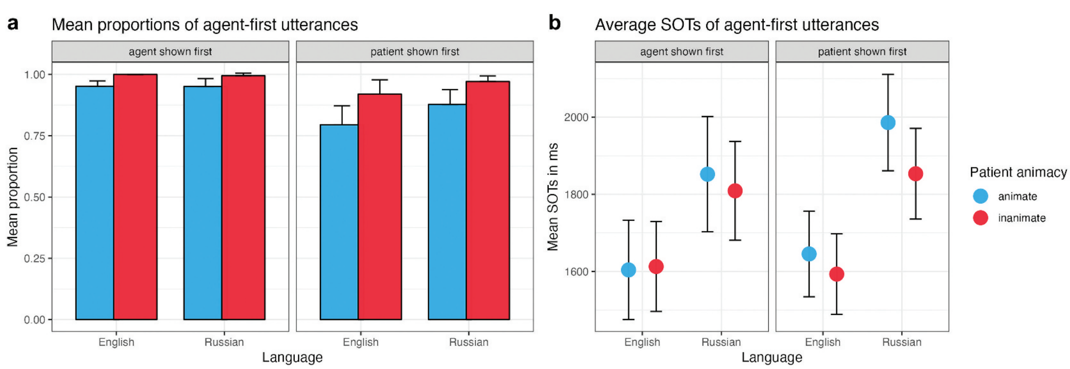

Figure 2. By-participant proportions of agent-first responses and speech onset times.

number as an additional (centered) predictor, with an interaction with Language, to test for the possibility of speakers adjusting their strategies during the course of the experiment.

As the degree to which speakers engage in incremental encoding is at least to some extent under their control and can change over time (e.g., Ferreira & Swets, 2002; Konopka et al., 2018; Wagner et al., 2010), participants may have adapted to the preview manipulation and to the type of stimuli they encountered after gaining familiarity with the task (note that our preregistered predictions did not include this analysis). Model comparisons showed a better fit for the second model (elpd_diff = −5.0, se_diff = 0.7), and because we believe that these results provide more insight into the data, we only report the results from the second model given space limitations. Importantly, the results for our main variables of interest (Preview Condition, Patient Animacy and Language) are very similar in both models (see supplementary materials for details on both models, including model comparisons, prior summaries, and postpredictive checks).

Table 2 reports the median and 95% credible interval (CI) to describe the posterior distributions of sampled regression weights, point direction, percent in ROPE (region of practical equivalence), Bayes factors (BF), and ESS (effective sample size). The *bayestestR* package (Makowski et al., 2019a) was used to retrieve these indices from all models. We take point direction to indicate the probability of the direction of an effect, and percent of HDI (Highest Density Interval) within ROPE as a measure of precision (Kruschke, 2014). We use the recommended default ROPE range for logistic regression models of – /+0.18. As there is no general agreement regarding the different indices of significance that can be obtained within a Bayesian modeling approach, we followed the suggestion of Makowski et al. (2019b), according to which interpretation can be based on the percentage of HDI in ROPE, viewed as a continuous index of significance. We define 5% for significance and 95% for nonsignificance as cutoff points. In other words, if less than 5% of the HDI is in the ROPE, we consider an effect as meaningful for our interpretation.

Given the data and the model, the effects of individual predictors were generally consistent with the hypotheses. The effect of Preview Condition had a probability of 99.62% of being positive (Median = 1.45, 95% CI [0.38, 2.55]). 0% of the HDI was in the ROPE, suggesting strong support for the hypothesis that more agent-first responses would be produced in the agent-preview than in the patient-preview condition. The effect of Patient Animacy had a probability of 99.82% of being positive (Median = 1.59, 95% CI [0.54, 2.64]). 0% of the HDI was in the ROPE, suggesting strong support for the hypothesis that more agent-first responses would be produced when the patient is inanimate than when it is animate. Importantly, the interaction between Preview Condition and Patient Animacy had a probability of 58,74% of being negative (Median = – 0.15, 95% CI [−1.46, 1.17]). With 22.28% of the HDI being in the ROPE, we interpret the effect as unreliable. Thus, although speakers produced numerically fewer agent-first descriptions when the previewed patient was animate than inanimate, early availability of information about the patient did not systematically modulate starting point choices.

There were also no strong differences in production choices between the groups of speakers. English speakers were numerically less likely to produce agent-first responses than Russian speakers (particularly

Table 2. Summary of posterior distribution.

Parameter	Median	95% CI	pd	% in ROPE	BF	ESS
Intercept	4.72	[3.99, 5.53]	100%	0%	7.91e+14	30,817.00
Preview Condition	1.45	[0.38, 2.55]	99.62%	0%	20.25	34,101.00
Patient Animacy	1.59	[0.54, 2.64]	99.82%	0%	40.36	39,740.00
Language	0.58	[−0.39, 1.57]	88.47%	15.14%	1.00	36,944.00
Trial Number	0.22	[−0.02, 0.47]	96.07%	37.65%	0.585	68,306.00
Preview Condition:Patient Animacy	−0.15	[−1.46, 1.17]	58.74%	22.28%	0.687	62,419.00
Preview Condition:Language	−0.85	[−2.24, 0.56]	88.22%	10.13%	1.47	53,200.00
Patient Animacy:Language	−0.74	[−1.90, 0.40]	90.16%	11.33%	1.33	59,353.00
Language1:Trial Number	0.76	[0.28, 1.25]	99.89%	0%	27.27	66,618.00
Preview Condition:Patient Animacy:Language	0.46	[−1.14, 2.05]	71.65%	15.89%	0.959	67,675.00

ROPE is [−0.18, 0.18]; all Rhat values ~1.00.

when an animate patient was previewed), but the effect of Language as well as all interactions with Language were not reliable (at least 10.13% of the HDI for these effects was in the ROPE; see Table 2).

Interestingly, the analysis showed a different pattern of responses in the two groups of speakers over time. The effect of Trial Number by itself had a probability of 96.07% of being positive (Median = 0.22, 95% CI [−0.02, 0.47]), but with 37.65% of the HDI being in the ROPE, we interpret the effect as unreliable. However, the slope difference captured by the interaction term for Language and Trial Number has a probability of 99.89% of being positive (Median = 0.76, 95% CI [0.28, 1.25]). 0% of the HDI was in the ROPE, suggesting strong support for the reliability of this effect. Figure 3 shows that exposure to and familiarity with the experiment had a different effect on speakers of English and Russian: English speakers showed a tendency to produce fewer agent-first sentences (and thus more patient-first responses) as the experiment went on, while the opposite was true of Russian speakers.

Speech onset times

Due to a small number of patient-first responses, we compared SOTs across conditions and language groups only for agent-first responses. We used the *brms* R package to set up Bayesian linear mixed models assuming a log-normal distribution. The trunc-function was used to specify that 4000 ms were the longest responses that were allowed. The levels of all predictors were coded in the same way as in the model reported here. Random effects terms for subjects and items were specified in the same way and for the same reasons as in the analysis reported.[8] Parameters were fitted again with 4 MCMC chains as described here.

The overall intercept had a student-*t* prior specifying a mean of 7.4 (estimated by the get_prior function) and a SD of 2.5 (student_t(3, 7,4, 2.5). The \sigma prior was specified as (gamma(1, 1)). Regression weights had normally distributed priors (normal(0, 1)), standard deviations of random effects had student t distributed priors, $\sigma \sim$ student_t(3, 0, 2.5), and the covariance matrices had LKJ Cholesky hyperpriors with $\eta = 1$. Model convergence and model fit was evaluated as in the previous analyses. We again set up a second model with Trial Number as a predictor and tested for an interaction with Language. Model comparison revealed a better fit for this model, and we report the results of this model below (see supplementary materials for details on both models). We use the recommended default ROPE range for linear regression models of ±0.1. Table 3 shows the results.

The effect of Preview Condition had a probability of 100% of being negative (Median = – 0.06, 95% CI [−0.09, – 0.04]). 0% of the HDI was in the ROPE, suggesting strong support for the hypothesis that agent-first responses had shorter SOTs in the agent-preview than in the patient-preview condition.

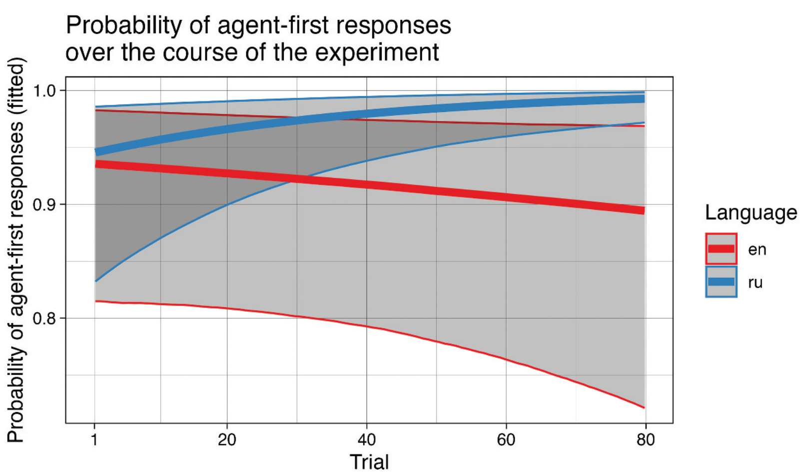

Figure 3. Probability of English and Russian agent-first responses produced over the course of the experiment (sampled from the posterior distribution).

Table 3. Summary of posterior distribution from the analysis of speech onset times (SOTs).

Parameter	Median	95% CI	pd	% in ROPE	BF	ESS
Intercept	7.43	[7.37, 7.48]	100%	0%	Inf	2936.00
Preview Condition	−0.06	[−0.09, −0.04]	100%	0%	314.30	12,673.00
Patient Animacy	−0.03	[−0.10, 0.03]	85.67%	15.24%	0.055	5832.00
Language	0.15	[0.06, 0.24]	99.91%	0%	8.26	2817.00
Trial Number	0.01	[0.00, 0.02]	97.28%	58.75%	0.030	67,143.00
Preview Condition:Patient Animacy	0.04	[0.00, 0.09]	97.02%	4.51%	0.150	25,667.00
Preview Condition: Language	−0.01	[0.05, 0.04]	66.09%	32.30%	0.025	13,037.00
Patient Animacy:Language	−0.02	[−0.06, 0.03]	78.09%	27.10%	0.029	28,803.00
Language:Trial Number	0.01	[−0.01, 0.03]	92.22%	35.52%	0.026	51,151.00
Preview Condition:Patient Animacy:Language	−0.02	[−0.10, 0.06]	67.44%	18.37%	0.045	28,838.00

ROPE is [−0.01, 0.01]; all Rhat values ~1.00.

This is consistent with the results of the starting points analysis and lends additional support to the validity of the preview manipulation: earlier SOTs are expected under the assumption that speakers started the sentence planning process by encoding information specific to the first-shown referent.

The analysis also showed a theoretically interesting effect of patient animacy. Numerically, there was a tendency for longer SOTs for responses in the animate patient condition. The effect of Patient Animacy had a probability of 85.67% of being negative (Median = – 0.03, 95% CI [–0.10, 0.03]), but with 15.24% of the HDI in ROPE, we consider the effect as unreliable. Importantly, however, the interaction between Patient Animacy and Preview Condition had a probability of 97.02% of being positive (Median = 0.04, 95% CI [0.00, 0.09]). With 4.51% of the HDI being in the ROPE, we take the effect as meaningful: the slope difference between animate and inanimate patients differed across the two preview conditions. Specifically, in the agent-preview condition, there was little difference in the SOTs of descriptions with animate and inanimate patients, while in the patient-preview condition, SOTs were longer in agent-first descriptions of scenes with animate than inanimate patients. Note that we did not observe an interaction between Preview Condition and Patient Animacy in the analysis of starting points, and we return to this point in the Discussion.

The effect of Language had a probability of 99.91% of being positive (Median = 0.15, 95% CI [0.06, 0.24]), and with 0% of the HDI in the ROPE, this suggests strong support for the hypothesis that SOTs were longer in Russian than English responses. This seems to be a general language effect, as the model shows no support for any interactions between Language, Preview Condition and Patient Animacy (all effects had at least 18% of the HDI in the ROPE).

The effect of Trial Number had a probability of 97.28% of being positive (Median = 0.01, 95% CI [0.00, 0.02]), as numerically, speakers become slower to respond over the course of the experiment. However, with 58.75% of the HDI in ROPE, we consider the effect as unreliable. Similarly, the interaction between Trial number and Language, with a probability of 92.22% of being positive (Median = 0.01, 95% CI [–0.01, 0.03]), and 35.52% of the HDI in the ROPE, cannot be interpreted as reliable. Thus, while experience with the task affected speakers' responses differently in the two languages, it had no effect on sentence planning as reflected in SOTs.

Exploratory analysis of excluded responses

Given that a large number of responses were not scorable and had to be excluded from the analyses reported above, we also investigated whether excluded trials were distributed equally across conditions and languages. An unequal distribution may suggest differences in encoding demands. An exploratory analysis was performed on the trials that remained after the removal of participants with fewer than 50% scorable data and after the removal of the items with fewer than 50% valid responses (see above). The dependent variable was whether a trial was excluded (binary-coded; yes = 1, no = 0). We used the *brms* R package to set up Bayesian generalized mixed models assuming a Bernoulli distribution with a logit link function[9] (See supplementary materials for details and the full model output).

Two effects appeared to be reliable, given the data and the model. The effect of Patient Animacy had a probability 98.67% of being negative (Median = −0.61, 95% CI [−1.15, −0.08]), and with 3.24% of the HDI being in the ROPE, we interpret the effect as meaningful: more responses on trials with animate patients had to be excluded relative to responses on trials with inanimate patients. Interestingly, as indicated by the effect of the interaction between Language and Patient Animacy, which had a 99.98% probability of being negative (Median = −1.30, 95% CI [−1.92, −0.66], with 0% of HDI in the ROPE), more responses on trials with animate patients had to be discarded in Russian than in English. This analysis may suggest that Russian speakers experienced more difficulty with scenes that showed animate patients.

Discussion

Previous research showed that speakers prefer to encode agents before patients when describing pictures of transitive events, but also that increasing the accessibility of patient referents can significantly reduce the agent-first bias, resulting in the production of more passive utterances (PVA in English) and more active utterances with the patient in sentence-initial position (PVA/PAV in Russian; Myachykov & Tomlin, 2008; Pokhoday et al., 2019). Here, we manipulated the order in which information about referents varying in animacy became available by providing a brief preview of either the agent or patient. We assessed how our manipulations influence the selection of starting points and speech onset times in English and Russian. The results showed three key findings. The *first* finding relates to the effects of the experimental manipulations on the selection of starting points (i.e., production of agent-first vs. patient-first sentences) and SOTs for agent-first sentences, the *second* concerns cross-linguistic differences, and the *third* concerns changes in encoding strategies over the course of the experiment.

First, the early availability of the patient referent and patient animacy modulated speakers' production choices. Speakers' overall preference for agent-first sentences was reduced when the patient referent was shown prior to the full scene. Speakers also produced fewer agent-first and more patient-first sentences when the patient referent was animate than inanimate. Interestingly, patient animacy and preview condition did not interact significantly: previewing an animate patient did not reliably increase the production of patient-first sentences compared to previewing an inanimate patient (although a numerical increase in this direction was observed).

The independence of the preview manipulation and the patient animacy manipulation shows that previewed animate characters were not automatically more likely to become starting points. Thus, speakers did not follow a strictly linear encoding strategy, as if event structure information had no influence on starting point choices. Instead, this effect suggests that the animacy of a previewed referent (i.e., its initially increased accessibility) was weighed against its thematic role which became apparent after full scene onset. In other words, when a previewed character turned out to be a patient on full scene availability, this fact contributed to speakers' decision to produce a patient-initial sentence more than the animacy of the patient. The finding that relational information processing triggered by perceiving the full scene affected encoding choices, taken on its own, is in line with the predictions of Hierarchical Incrementality.

This interpretation based on the starting point analysis alone, however, is challenged by the SOTs of agent-first responses. SOTs were longer in the patient-preview than in the agent-preview condition, and in contrast to the analysis of starting points, the SOT analysis showed an interaction of the preview manipulation with patient animacy. In the agent-preview condition, SOTs did not differ significantly between items with animate and inanimate patients. Thus, when the previewed referent was deemed to be a "good" starting point during the earliest stages of encoding (i.e., when it was animate), the animacy of the not-previewed referent (the patient) did not influence the onset of articulation significantly.

The crucial finding, however, is that, animacy influenced SOTs if it was a feature of the previewed patient referent: in the patient-preview condition, SOTs (of agent-first responses) were longer in sentences with animate patients than inanimate patients. This suggests that speakers consistently started the planning process by encoding the previewed animate referent with priority, possibly on the assumption that an animate referent would turn out to be an agent (see below for an alternative interpretation). When the first-shown animate referent in the animate patient preview condition then turned out to be a patient, speakers may have been forced to revise their initial message and sentence plan, resulting in longer SOTs. Put differently, the referent that eventually ended up in sentence-initial position (the agent in most cases; see the starting point analysis) was not always the referent that speakers began to encode first. Importantly, these effects indicate that the early availability of visual information influenced early processing: speakers did start the encoding process with information specific to the referent that was shown first, consistent with Linear Incrementality. This revision process is a novel demonstration of the costs and benefits of linearly incremental planning.

The second finding of interest is that processes involved in the selection of starting points differed across the two language groups. Russian speakers were numerically more likely to produce agent-first responses overall, and previewing a patient referent resulted in a numerically smaller drop in agent-first responses than in English speakers. This is consistent with the findings of Myachykov and Tomlin (2008), who argued that the relatively small effects of changes in patient accessibility on the selection of starting points was likely due to the low frequency of noncanonical structures in Russian and/or to the fact that production of patient-first structures was motivated more strongly by discourse than by perceptual salience. Note that the frequency of patient-first structures we observed for Russian speakers is lower than in previous studies (e.g., Myachykov & Tomlin, 2008). This could be due to the timing of the stimulus presentation in the current study. More specifically, the preview duration in combination with the immediately following onset of the full scene might have created a situation that tapped into a very early phase during grammatical encoding during which speakers are still more prone to revise initial encoding plans.

An effect of language was observed in the SOT analysis for agent-first sentences, too. Russian speakers had longer SOTs than English speakers across all conditions. This is consistent with Myachykov et al. (2013), where a similar language difference was explained by the greater syntactic flexibility offered by Russian grammar, which leads to greater competition during the initial stages of the syntactic encoding process. An alternative explanation could be that Russian speakers were more likely to follow a "wait-and-see"-strategy, i.e., they were less willing than English speakers to start planning sentences on the basis of partial information. However, our finding that SOTs for agent-first responses strongly differed between animate and inanimate patient-preview trials in Russian, whereas they hardly differed in the two agent-preview conditions renders such an explanation unlikely. Instead, this finding suggests that Russian speakers did commit to early encoding choices, similar to English speakers. This leaves the larger number of structural alternatives in Russian as a better explanation of the overall difference in SOTs between English and Russian. Importantly, the SOT difference between animate and inanimate patient preview trials also shows evidence of linear incrementality in Russian.

Russian speakers also differed from English speakers in that they produced more invalid responses on animate-patient trials, which indicates that they encountered more encoding difficulties whenever there was a conflict between the animacy of the referent and its suitability for a starting point, which could not be resolved successfully by a revision in the duration of an experimental trial.

The third key finding concerns how starting point choices change over the course of the experiment: production of agent-first descriptions dropped in English but increased in Russian. Thus, speakers of the two languages adapted to the experiment in different ways: English speakers lowered their initial encoding bias to place the agent in sentence-initial position, while Russian speakers showed an increasingly stronger bias to place the agent in sentence-initial position.

Next, we discuss our results in light of more general encoding principles that, according to the Production-Distribution-Comprehension (PDC) account, can together determine the extent to which speakers may engage in linear incremental encoding: (1) Easy First, (2) Plan Reuse, and (3) Reduce Interference (MacDonald, 1999, 2013). By decomposing the more specific theoretical assumptions of linear vs. hierarchical incremental encoding into these underlying computational principles, we can address what might otherwise be considered alternative explanations of the findings (such as self-priming). According to *Easy First*, information that is more accessible (and available earlier for linguistic processing) is encoded before information that is less accessible (and available later). *Plan Reuse* captures that speakers show a bias to reuse previously employed plans (global goal and subgoal plans). *Reduce interference* helps to prevent interference resulting from competition between simultaneously available elements. The three principles can jointly mitigate encoding difficulty and thus may pull encoding choices in different directions.

In our experiment, *Easy First* and *Reduce Interference* both push speakers toward attempting to produce utterances by encoding the previewed referent as early as possible, because already encoded elements cannot interfere with not-yet encoded elements. Earlier encoding of some element also frees up resources for the encoding of subsequent elements. However, committing too early to producing a given referent in sentence-initial position can be costly and may require a time-consuming revision. *Plan Reuse* can relate to a global goal (e.g., encoding a whole sentence) as well as a subgoal (e.g., encoding a local phrase). If *Plan Reuse* concerns the global plan, one would expect "classic" structural priming effects, such as the repetition of active syntax from one trial to the next (e.g., Bock, 1986). If *Plan Reuse* concerns a subgoal, e.g., the encoding of a noun phrase in sentence-initial position, one might expect the reuse of the same noun phrase structure as in previous trials, including syntactic specifications (e.g., determiner, subject function, and importantly, *nominative case*). Given that we observed differences between English and Russian, it appears that these principles are weighted differently in the two languages, both with respect to their default settings and also in terms of the way that these weights can change over time. It is possible that Russian speakers assign global *Plan Reuse* a higher weight than English speakers, resulting in a higher degree of persistence or repetition of the same syntactic specifications of sentence-initial NPs across sentences.

With respect to the *Reduce Interference* principle, we start by discussing how far our data show evidence of interference during encoding. English speakers produced *reduced* utterances which omitted auxiliaries (e.g., "telegraphic" utterances such as "A farmer filming a fisherman" as opposed to "A farmer is filming a fisherman") on more than 68% of all critical trials. Although this was not expected initially, it allows us to draw inferences regarding the extent to which interference arose during production of English utterances. If English speakers were affected by interference, we would expect a high frequency of unreduced sentences, because such structures allow for the greatest possible distance between the two referents, consequently reducing interference during encoding (see Gennari et al., 2012, for an in-depth discussion of relative clauses in relation to the *Reduce Interference* principle). Thus, the overall high frequency of *reduced* utterances suggests that English speakers did not experience high levels of interference. This can be interpreted as showing that English speakers did encode the previewed referent early (assigning it the subject function), and that increasing the temporal and syntactic distance between referents by introducing additional linguistic material between the two referents was not required.

Regarding interference in Russian, our data do not allow us to draw strong inferences. However, three observations may be relevant: Russian speakers did not produce telegraphic style structures comparable to those observed in English, SOTs were overall significantly longer compared to SOTs in English, and SOTs were longer when the stimulus showed more similar referents (animate patient condition). While overall prolonged SOTs in Russian compared to English have previously been interpreted to reflect higher cognitive load due to the selection of simultaneously activated syntactic frames (Myachykov et al., 2013), the effect might at least in part also be due to competition or interference between referents during function assignment, as previously discussed for Korean (Hwang & Kaiser, 2015). This issue could be investigated by comparing productions in scene descriptions with similar vs. dissimilar referents under a previewed and an unpreviewed condition.

Next, we discuss evidence for the influence of the *Plan Reuse* principle. As mentioned above, it is necessary to distinguish between the reuse of a syntactic frame (e.g., active vs. passive or double-object vs. prepositional object construction) and the reuse of local structures (e.g., a specific structure for an NP). As in previous studies, speakers of both English and Russian in our experiment showed a strong bias for reusing specific syntactic structures over and over again, suggesting that they relied on the *Plan Reuse* principle for global structures. Those English speakers who initially produced active reduced structures, stuck to this choice, breaking the pattern occasionally by producing passive reduced structures. Those English speakers who initially started with full declarative active sentences also stuck to this choice, producing full passives instead only occasionally. Active SVO structures (i.e., AVP sentences) were dominant in Russian, with OVS (or PVA) the most common structure for patient-first responses.

Importantly, given the specific manipulation in our study, we were also able to obtain independent evidence of speakers relying on the *Plan Reuse* principle for local structures, more specifically, the reuse of noun phrases that are assigned the subject function. Importantly, while syntactic subjects in English (in contrast to Russian) do not carry any overt marking, they still reveal their syntactic function in that they show agreement with the main verb (3rd person singular in our experiment). In other words, in English, two NPs that look identical on the surface may differ syntactically after they underwent syntactic encoding and have been assigned different syntactic functions (e.g., subject or direct object). Our results are compatible with the assumption that speakers showed a bias toward reusing subject NPs for animate previewed referents, which played out as follows: the first available noun lemma (previewed animate referent) is encoded as a syntactic subject due to *Plan Reuse* for local structures. In contrast to agent-preview trials, where *Plan Reuse* for global and local structures converge on the same structures (active SVO), applying both principles in the same fashion on animate patient-preview trials results in a dead end, because the patient cannot be the subject of an active SVO sentence. This assumption is supported by our finding that there was a significant difference in SOTs of agent-first responses between the animate and inanimate patient-preview conditions in both languages. Thus, if global *Plan Reuse* is weighted relatively high, speakers cannot proceed with their initial choice, and are required to revise their initial plan (or resort to producing an invalid structure). If *Plan Reuse* for global structures is weighted relatively lower than *Plan Reuse* for local structures, speakers can stick to their initial local structure – the first NP as the syntactic subject – and produce a nonactive sentence with the patient as the subject in first position. Our results show that English speakers were successful in overcoming the pressure from global *Plan Reuse* more often than Russian speakers. In line with this explanation, we also found more invalid trials in the animate patient-preview condition in Russian.

The interplay between the local and global *Plan Reuse* principles can explain the prolonged SOTs on animate patient-preview trials without having to assume that speakers assign the agent role to previewed animate referents by default, as hypothesized earlier: Previewed animate referents receive the highest syntactic function (subject) and are planned to be produced in sentence-initial position, as dictated by local *Plan Reuse*. At the same time, an active sentence is initially prepared, which requires an agent in sentence-initial position. Upon seeing the whole scene, the role of the previewed referent is revealed (i.e., the fact that this referent is a patient, not an agent), which makes it impossible to continue with the encoding plan established so far. The previewed referent is reassigned the object function, and the second referent receives the subject function, allowing the production of an agent-first sentence, as favored by global *Plan Reuse*. This revision process takes time. According to this interpretation, syntactic encoding may be initiated without any conceptual planning in our experiment.

To conclude, the study reported here collected data in a web-based language production task where sentences were produced out of context. Production could not be monitored during the experimental sessions and thus participants were not reminded to avoid producing unscorable or incomplete sentences, to provide answers more quickly, or to try to avoid errors. Yet, the responses produced by participants, the errors that they made, and the speech onset times observed in scorable responses provided valuable insight into questions regarding the extent to which speakers of different languages are able and willing to use information that is presented in a piecemeal fashion during sentence planning. In this regard, the present study is a continuation of the line of work for which Russel Tomlin is well known in psycholinguistics.

Notes

1. Slioussar and Makarchuk (2022) give the following frequencies of these structures for a corpus consisting of subtitles, which comes closest to spoken language (see also Myachykov et al., 2013): SVO (63.3%), SOV (18.2%), OSV (14.4%), OVS (2.5%), VOS (1.0%), VSO (0.6%).
2. We note that more syntactic options are possible in Russian than are included in this list. The additional options were deemed unlikely in the current study, according to feedback from a native speaker of Russian. Although Russian allows different word order combinations for the target sentences elicited in the current study, the choice between these word order options and syntactic structures is largely driven by information structure and by stylistic concerns. Given that participants produced single-sentence descriptions out of context, this effectively reduced the number of *plausible* options. A wider range of structures would be expected in discourse-rich contexts.
3. Russian speakers use the perfective form of the verb and past tense for punctual events (or verbs) like "to sting."
4. There were also 4 responses in the Russian data set that had the verb in first position. These were also excluded, because they could not reliably be assigned to either of the two relevant response categories (agent-first or patient-first sentences).
5. There were 3 occurrences of passive structures omitting the agent in English, and 4 occurrences in Russian. We included these structures in the analysis, on the reasoning that they still demonstrate a speaker's commitment to a specific starting point. Removing these sentences from the analysis does not change the outcome or interpretation of the effects of interest.
6. The dataset included only 2% non-subject-first sentences (0% in English, 4% Russian: 54 responses, 40 of which were in the patient-preview condition), so we omitted an analysis of syntactic function assignment.
7. bf(response ~ 1 + condition * animacy * language + (1 + condition * animacy|subject) + (1 + language * condition|item), family = bernoulli(link = "logit"))
8. bf(SOT | trunc(ub = 4000) ~ 1 + condition * animacy * language + (1 + condition * animacy|subject) + (1 + condition * language |item), family = lognormal(link = "identity")
9. bf(excluded ~ 1 + condition * animacy * language + (1 + condition * animacy| subject) + (1 + condition * language |item), family = bernoulli(link = "logit"))

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

Parts of this work were supported by the German Research Foundation, Project-ID 281511265 – as part of the CRC 1252 "Prominence in Language" in project B06 "Attention and prominence in language production" at the University of Cologne.

Data availability statement

The data and analysis code that support the findings of this study are openly available on the Open Science Framework (OSF) at https://osf.io/yxat5/.

This study was pre-registered on OSF at https://osf.io/th5y6.

References

Bock, J. K., & Warren, R. K. (1985). Conceptual accessibility and syntactic structure in sentence formulation. *Cognition*, *21*(1), 47–67. https://doi.org/10.1016/0010-0277(85)90023-X

Bock, K., & Ferreira, V. S. (2014). Syntactically speaking. In M. Goldrick, V. Ferreira, & M. Miozzo (Eds.), *The Oxford handbook of language production* (pp. 21–46). Oxford University Press.

Bock, K., & Griffin, Z. M. (2000). The persistence of structural priming: Transient activation or implicit learning? *Journal of Experimental Psychology: General*, *129*(2), 177. https://doi.org/10.1037/0096-3445.129.2.177

Bock, K., Irwin, D., & Davidson, D. J. (2004). Putting first things first. In J. M. Henderson & F. Ferreira (Eds.), *The Interface of language, vision, and action* (pp. 249–278). Psychology Press.

Bock, K. J. (1986). Syntactic persistence in language production. *Cognitive Psychology*, *18*(3), 355–387. https://doi.org/10.1016/0010-0285(86)90004-6

Bock, K., & Levelt, W. (1994). Grammatical Encoding. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. S. 945–984). Academic Press.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International*, *5*(9/10), 341–345.

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2009). The role of prominence information in the real-time comprehension of transitive constructions: A cross-linguistic approach. *Language and Linguistics Compass*, *3*(1), 19–58. https://doi.org/10.1111/j.1749-818X.2008.00099.x

Branigan, H. P., Pickering, M. J., & Tanaka, M. (2008). Contributions of animacy to grammatical function assignment and word order during production. *Lingua*, *118*(2), 172–189. https://doi.org/10.1016/j.lingua.2007.02.003

Bürkner, P. C. (2017). brms: An R package for Bayesian multilevel models using stan. *Journal of Statistical Software*, *80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Christianson, K., & Ferreira, F. (2005). Conceptual accessibility and sentence production in a free word order language (Odawa). *Cognition*, *98*(2), 105–135. https://doi.org/10.1016/j.cognition.2004.10.006

Dolscheid, S., & Penke, M. 2023. Attention vs. accessibility: The role of different cue types for non-canonical sentence production in German. *Frontiers in Language Sciences*, *2*, 1256471.

Duñabeitia, J. A., Crepaldi, D., Meyer, A. S., New, B., Pliatsikas, C., Smolka, E., & Brysbaert, M. (2018). MultiPic: A standardized set of 750 drawings with norms for six European languages. *Quarterly Journal of Experimental Psychology*, *71*(4), 808–816. https://doi.org/10.1080/17470218.2017.1310261

Egurtzegi, A., Blasi, D. E., Bornkessel-Schlesewsky, I., Laka, I., Meyer, M., Bickel, B., & Sauppe, S. (2022). Cross-linguistic differences in case marking shape neural power dynamics and gaze behavior during sentence planning. *Brain and Language*, *230*, 105127. https://doi.org/10.1016/j.bandl.2022.105127

Esaulova, Y., Dolscheid, S., & Penke, M. (2021). All it takes to produce passives in German. In V. Torrens (Ed.), *Syntax processing* (pp. 75–107). Cambridge Scholars Publishing.

Esaulova, Y., Penke, M., & Dolscheid, S. (2019). Describing events: Changes in eye movements and language production due to visual and conceptual properties of scenes. *Frontiers in Psychology*, *10*, 1–15. https://doi.org/10.3389/fpsyg.2019.00835

Esaulova, Y., Penke, M., & Dolscheid, S. (2020). Referent cueing, position, and animacy as accessibility factors in visually situated sentence production. *Frontiers in Psychology*, 11, 1–13. https://doi.org/10.3389/fpsyg.2020.02111

Ferreira, F., & Swets, B. (2002). How incremental is language production? Evidence from the production of utterances requiring the computation of arithmetic sums. *Journal of Memory and Language*, *46*(1), 57–84. https://doi.org/10.1006/jmla.2001.2797

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*(4), 457–472. https://doi.org/10.1214/ss/1177011136

Gennari, S. P., Mirković, J., & MacDonald, M. C. (2012). Animacy and competition in relative clause production: A cross-linguistic investigation. *Cognitive Psychology*, *65*(2), 141–176.

Gleitman, L. R., January, D., Nappa, R., & Trueswell, J. C. (2007). On the give and take between event apprehension and utterance formulation. *Journal of Memory and Language*, *57*(4), 544–569. https://doi.org/10.1016/j.jml.2007.01.007

Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science*, *11*(4), 274–279.

Hafri, A., Papafragou, A., & Trueswell, J. C. (2013). Getting the gist of events: Recognition of two-participant actions from brief displays. *Journal of Experimental Psychology: General*, *142*(3), 880–905. https://doi.org/10.1037/a0030045

Hwang, H., & Kaiser, E. (2014). Having a syntactic choice is not always better: The effects of syntactic flexibility on Korean production. *Language, Cognition and Neuroscience*, *29*(9), 1115–1131. https://doi.org/10.1080/23273798.2013.875212

Hwang, H., & Kaiser, E. (2015). Accessibility effects on production vary cross-linguistically: Evidence from English and Korean. *Journal of Memory and Language*, *84*, 190–204. https://doi.org/10.1016/j.jml.2015.06.004

Indefrey, P., & Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition*, *92*(1/2), 101–144. https://doi.org/10.1016/j.cognition.2002.06.001

Konopka, A. E., & Brown-Schmidt, S. (2014). Message encoding. In V. Ferreira, M. Goldrick, & M. Miozzo (Eds.), *The Oxford handbook of language production* (pp. 3–20). Oxford University Press.

Konopka, A. E., Meyer, A., & Forest, T. A. (2018). Planning to speak in L1 and L2. *Cognitive Psychology*, *102*, 72–104. https://doi.org/10.1016/j.cogpsych.2017.12.003

Konopka, A. E., & Meyer, A. S. (2014). Priming sentence planning. *Cognitive Psychology*, *73*, 1–40.

Kruschke, J. (2014). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan* (2nd ed.). Academic Press.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. MIT Press.

MacDonald, M. C. (1999). Distributional information in language comprehension, production, and acquisition: three puzzles and a moral. In ed B. MacWhinney," in *The Emergence of Language* (pp. 177–196). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology*, *4*, 226.

Makowski, D., Ben-Shachar, M., & Lüdecke, D. (2019a). bayestestR: Describing effects and their uncertainty, existence and significance within the Bayesian framework. *Journal of Open Source Software*, *4*(40), 1541. https://doi.org/10.21105/joss.01541

Makowski, D., Ben-Shachar, M. S., Chen, S. A., & Lüdecke, D. (2019b). Indices of effect existence and significance in the Bayesian framework. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.02767

Momma, S., Slevc, L. R., & Phillips, C. (2016). The timing of verb selection in Japanese sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(5), 813–824. https://doi.org/10.1037/xlm0000195

Myachykov, A., Garrod, S., & Scheepers, C. (2018). Attention and memory play different roles in syntactic choice during sentence production. *Discourse Processes*, *55*(2), 218–229. https://doi.org/10.1080/0163853X.2017.1330044

Myachykov, A., Garrod, S., & Scheepers, C. (2011). Perceptual priming of structural choice during English and Finnish sentence production. In R. K. Mishra & N. Srinivasan (Eds.), *Language-Cognition interface: State of the Art* (pp. 53–71). LINCOM Europa.

Myachykov, A., Scheepers, C., Garrod, S., Thompson, D., & Fedorova, O. (2013). Syntactic flexibility and competition in sentence production: The case of English and Russian. *The Quarterly Journal of Experimental Psychology*, *66*(8), 1601–1619. https://doi.org/10.1080/17470218.2012.754910

Myachykov, A., Thompson, D., Garrod, S., & Scheepers, C. (2012). Referential and visual cues to structural choice in visually situated sentence production. *Frontiers in Psychology*, 2,1–9. https://doi.org/10.3389/fpsyg.2011.00396

Myachykov, A., & Tomlin, R. (2008). Perceptual priming and structural choice in Russian sentence production. *Journal of Cognitive Science*, *9*(1), 31–48. https://doi.org/10.17791/jcs.2008.9.1.31

Norcliffe, E., Konopka, A. E., Brown, P., & Levinson, S. C. (2015). Word order affects the time course of sentence formulation in Tzeltal. *Language, Cognition and Neuroscience*, *30*(9), 1187–1208. https://doi.org/10.1080/23273798.2015.1006238

Pokhoday, M., Shtyrov, Y., & Myachykov, A. (2019). Effects of visual priming and event orientation on word order choice in Russian sentence production. *Frontiers in Psychology*, *10*, 1–8. https://doi.org/10.3389/fpsyg.2019.01661

Prat-Sala, M., & Branigan, H. P. (2000). Discourse constraints on syntactic processing in language production: A cross-linguistic study in English and Spanish. *Journal of Memory and Language*, *42*(2), 168–182. https://doi.org/10.1006/jmla.1999.2668

R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Sauppe, S., Norcliffe, E., Konopka, A. E., Van Valin, R. D. J., & Levinson, S. C. (2013). Dependencies first: Eye-tracking evidence from sentence production in Tagalog. In M. Knauff, M. Pauen, N. Sebanz, & E. Wachsmuth (Eds.), *Proceedings of the 35th annual meeting of the Cognitive Science Society* (pp. 1265–1270). Cognitive Science Society.

Schlenter, J., Esaulova, Y., Dolscheid, S., & Penke, M. (2022). Ambiguity in case marking does not affect the description of transitive events in German: Evidence from sentence production and eye-tracking. *Language, Cognition and Neuroscience*, *37*(7), 844–865. https://doi.org/10.1080/23273798.2022.2026419

Schlenter, J., & Penke, M. (2022). (How) Visual properties affect the perception and description of transitive events. *Glossa Psycholinguistics*, *1*(1), 5,1–29. https://doi.org/10.5070/G601193

Slioussar, N., & Makarchuk, I. (2022). SOV in Russian: A corpus study. *Journal of Slavic Linguistics*, *30*(3), 1–14.

Tomlin, R. (1986). *Basic word order: Functional principles*. Croom Helm.

Tomlin, R. S. (1995). Focal attention, voice, and word order. In P. Dowing & M. Noonan (Eds.), *Word order in discourse* (pp. 517–552). John Benjamins.

Tomlin, R. S. (1997). Mapping conceptual representations into linguistic representations: The role of attention in grammar. In J. Nuyts & E. Pederson (Eds.), *Language and conceptualization* (pp. 162–189). Cambridge University Press.

Van de Velde, M., Meyer, A. S., & Konopka, A. E. (2014). Message formulation and structural assembly: Describing "easy" and "hard" events with preferred and dispreferred syntactic structures. *Journal of Memory and Language*, *71*(1), 124–144.

Van Nice, K. Y., & Dietrich, R. (2003). Task sensitivity of animacy effects: Evidence from German picture descriptions. *Linguistics*, *41*, 825–849.

Wade, T. L. B. (2011). *A comprehensive Russian grammar*. Wiley-Blackwell.

Wagner, V., Jescheniak, J. D., & Schriefers, H. (2010). On the flexibility of grammatical advance planning during sentence production: Effects of cognitive load on multiple lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(2), 423–440. https://doi.org/10.1037/a0018619

Zehr, J., & Schwarz, F. (2018). *PennController for Internet Based Experiments (IBEX)*. https://doi.org/10.17605/OSF.IO/MD832