

Krasimir SPASOV*(American University in Bulgaria)*

STRUCTURING INFORMATION: INSIGHTS FROM A CORPUS-BASED STUDY OF PASSIVE VOICE

Abstract. *The present paper explores how speakers use the passive voice in spoken English as a practical tool for organizing information and guiding listeners through a message. Rather than being just a stylistic alternative to the active voice, the passive often helps speakers highlight what is already familiar, keep a topic in focus, leave out agents that do not matter, and place heavier or more complex details towards the end of a clause. Using examples from the Corpus of Contemporary American English (COCA), this paper shows that these choices reflect broader tendencies in the English language toward presenting information from given to new and maintaining a smooth flow of ideas. The findings suggest that the passive voice plays a meaningful role in real-time communication by helping speakers manage attention and process demands.*

Keywords: *passive voice; information packaging; information structure; corpus linguistics; spoken English; cognitive load*

Красимир СПАСОВ*(Американски университет в България)*

ИНФОРМАЦИОННА СТРУКТУРА: ПОГЛЕД ВЪРХУ СТРАДАТЕЛНИЯ ЗАЛОГ В ПРИМЕРИ ОТ КОРПУСА НА СЪВРЕМЕННИЯ АМЕРИКАНСКИ АНГЛИЙСКИ ЕЗИК

Резюме. *Настоящата статия изследва как говорещите използват пасивния залог в разговорния английски език като практичен инструмент за организиране на информацията и навигиране на слушателите в рамките на комуникацията. Пасивният залог не е просто стилистично маркиран вариант на активния, а по-скоро помага на говорещите да подчертаят това, което е вече познато, да запазят фокуса върху темата, да пропуснат деятелите,*

които нямат значение, както и да поставят по-тежките или по-сложни елементи в края на изречението. Базирана на примери от Корпуса на съвременния американски английски език (COCA), тази статия показва, че споменатите избори отразяват широко разпространени тенденции в английския език към представяне на информацията в посока от известна към нова и поддържане на постъпателно разгръщане на идеите. Резултатите показват, че пасивният залог играе значителна роля в комуникацията в реално време, като помага на говорещите да насочват вниманието и анализират търсенията на информация.

Ключови думи: пасивен залог; информационно пакетиране; информационна структура; корпусна лингвистика; разговорен английски език; когнитивна тежест

Introduction

Information packaging is one of the most fundamental aspects of human communication. It determines how speakers (and writers) structure their messages to effectively convey meaning while managing a balance between both cognitive load and discourse flow. At the intersection of syntax, semantics, and pragmatics lies the intricate relationship between information structure and choice of grammatical voice, especially in the case of the strategic use of passive voice constructions.

The passive voice has long been recognized as more than a mere syntactic alternative to active constructions – it serves as a sophisticated tool for information management that allows speakers to manipulate focus, topicality, and thematic progression. While traditional prescriptivist approaches have often discouraged passive voice usage (Strunk & White 2002: 18 – 19), contemporary linguistic research reveals its crucial role in organizing information according to discourse-functional principles such as the given-new information, end-weight, and thematic continuity (Birner & Ward 2009). This article examines the relationship between information packaging strategies and passive voice employment within the spoken context. I argue that passive voice selection is fundamentally driven by information-structural considerations, serving to:

- promote thematic subjects to sentence-initial position,
- omit agents when contextually irrelevant,
- maintain topic continuity across discourse segments
- achieve end-weight balance for information processing efficiency

Through the analysis of authentic discourse corpus data, this study demonstrates how passive voice functions as a grammaticalized information

packaging device. It aims to reveal the intricate ways in which syntactic choices reflect and facilitate cognitive processing loads.

Literature review

Corpus studies consistently show that passive constructions are far less frequent in spoken interaction, typically around 2% of the finite verbs, while they occur much more often in written registers such as news and academic prose, where they account for roughly 25%. In spoken communication, passive voice serves clear pragmatic purposes such as backgrounding agency and maintaining focus (Banks 2017: 1 – 9).

At the heart of information structure is the distinction between core meaning and how we build the sentence constituents. Both passive and active voice often share truth-conditional content, but they differ in how information is organized for the hearer i.e., what is presented as given or presupposed and what is foregrounded as the new, focused element. Classic thematic differentiation (theme-rheme) and cognitive approaches (Firbas 1992) agree on the idea that speakers arrange propositions to match hearers' working memory and expectations: a well-packaged sentence orders known material before new material and places syntactically "heavy" constituents late in the clause to reduce processing load (Huddleston & Pullum 2002).

The theory of information packaging as established by Lambrecht (1994) is a primary framework for understanding why speakers choose marked sentence forms (including passive voice) to manage speaker-hearer expectations. Lambrecht's notion that the focused element is the least predictable part of the proposition provides a link to grammatical form with assumptions about addressee knowledge and unpredictability. Presupposition research over the last decade (Asami & Sugawara 2023) has managed to build tools for diagnosing which parts of an utterance are treated as backgrounded vs. asserted and for testing how such content can be "projected." Large annotation projects and experimental investigations (often grouped under "projectivity" or "PROPRES/ProPres" initiatives) have shown that presuppositions are not uniformly categorical: their projectivity and recoverability depend on lexical triggers, embedding contexts, prosody, and discourse assumptions. These findings are directly relevant to how passive constructions may rely on (or introduce) presupposed material whose accessibility varies across audiences.

Further experimental and corpus work suggests that prosodic prominence and the informational status of referents can influence whether content is taken as presupposed or treated as newly asserted. This line of research provides concrete diagnostics one can use to determine when the

focus content of passive voice is likely to be “available” to an addressee (Roberts 2012). Since 2015 corpus and diachronic studies have also been deepening our understanding of the functional diversity of passives. Rather than serving a single discourse role, passive voice is often classified into functional subtypes (reporting, procedural, relational, resultative, etc.), and corpus work documents both genre-specific distributions and diachronic shifts, including evidence for a modest decline of passive use in some scholarly registers (Mair & Leech 2006; Hou & Smith 2018). These empirical taxonomies help us explain why speakers select passive constructions in contexts where emphasizing the event or result (rather than the agent) is pragmatically advantageous.

Evidence from historical syntax further supports the claim that information-structural considerations guide speakers’ grammatical choices. In her 2016 study of Old English and Old Icelandic derived orders at the Syntax-Information Structure Interface, Chankova invokes semantic and information-structural factors in an attempt to determine to what extent the general linearization principles (weight, definiteness, pronominality) can be affected by such factors. Regarding structures involving both Scrambling and Passivization, she argues that the studied patterns of word order variation in the Middlefield and/or the Prefield are derived through leftward movement of the targeted constituents which is not triggered by feature-checking requirements but is semantically and pragmatically effective (Chankova, 2016, pp. 75 – 78; 86 – 88). The information packaging effects found with similar Old English modified orders comprising two non-finite verbs are discussed in Chankova 2021 (pp. 119 – 121). Together, these findings reinforce the broader cross-linguistic tendency for information-structural factors to shape the realization of passive and related non-canonical word orders.

Three pragmatic categories routinely recur in general literature review on informational structure and thus form the operative vocabulary for most analyses (Lambrecht 2001):

- Pragmatic presupposition (old/given information) versus pragmatic assertion (new information);
- Topic (what the utterance is about) versus Focus (the unpredictable or informative element); and
- Activation (the accessibility status of referents in the hearer’s mind: active, semi-active, inactive).

These categories explain why, for example, speakers will often promote an object to a subject position (via passive) when that object is

discourse-old or easily activated. Therefore, this reserves the clause-final position for heavy or new material in line with end-weight and information-flow preferences.

Methodology and corpus

This study uses data from the Corpus of Contemporary American English (COCA), which is a large, balanced, genre-diverse corpus that enables systematic retrieval of passive constructions across different registers. All tokens were extracted using standard passive-voice search query “BE+vvn,” with concordances retrieved alongside two to three surrounding sentences to ensure adequate discourse context. The verb “be” is entered in uppercase to allow for variation in tense, person, and number, while the “vvn” tag is used to capture all past participle forms of the lexical verbs. This highlights one of the practical advantages of COCA compared to the British National Corpus (BNC) for instance. In the BNC, this particular annotation yields only two instances (one in spoken English and one in fiction), whereas COCA returns fourteen tokens distributed across several registers (ten spoken, two fiction, one news, and one academic).

A random sample was then taken across the so-called spoken genre. The quantitative component consists of texts extracted directly from COCA to examine associations between passive type and information-status category. These quantitative findings are complemented by qualitative discourse analysis of selected examples. Together, these methods allow the study to describe not only how often certain passive constructions appear, but also how they function as information-packaging strategies when the speaker anticipates a fully informed, partially informed, or entirely unknowledgeable audience.

Information packaging analysis

Before turning to the qualitative analysis, it is useful to sketch the general landscape that emerged from the corpus search. The COCA queries produced an effective set of passive constructions headed by high-frequency past participles. For instance, the sequence “be made” returned 26,089 tokens across registers, “be taken” appeared in 18,283 tokens, and even the comparatively less common “be blown” surfaced nearly 4,000 times (3,982 tokens precisely). These numbers themselves are revealing: they signal that passive voice is not a rare stylistic construction but a regular, productive resource that speakers lean upon when unfolding information in a given discourse. These distributional patterns also provide a starting point for examining how speakers use the passive voice to manage expectations about

what the audience knows (or does not yet know) at the moment of speaking. In this section, I focus on a subset of spoken examples drawn from this larger dataset to illustrate how different passive constructions (namely, “be made,” “be blown,” “be taken”) operate under conditions of a given focus.

The following excerpt from COCA is the first part of a news segment focused on the latest developments in the death of Jeffrey Epstein by August 2019. According to the news outlet, officials released the results of his autopsy and confirmed that he had died by hanging in his Manhattan jail cell the previous Saturday. Epstein had been awaiting trial on federal sex-trafficking charges involving minors. His death immediately raised questions, and both the FBI and the Department of Justice opened investigations after what were described as “serious irregularities” discovered at the jail:

The results of the autopsy released today said Epstein hanged himself in his Manhattan jail cell last Saturday. Epstein was awaiting trial on federal sex trafficking charges. The FBI and Justice Department are both investigating Epstein's death after -- quote -- "serious irregularities" were found at the jail. There are new revelations today about the Air Force's probe into sexual assault allegations made against President Trump's pick for the Pentagon's second highest military post. Air Force investigators determined there was insufficient evidence to prove Air Force General John Hyten had a -- quote -- "unprofessional relationship" with his close aide Army Colonel Kathryn Spletstoser. Hyten's polygraph test was also deemed to be inconclusive. **A separate report from the Defense Department's inspector general could be made public as early as next week.** Hyten has denied the assault claim. He faces a full Senate confirmation vote next month. Meanwhile, a new report from the State Department's inspector general has found politically motivated harassment at one of the department's top bureaus. Career staffers in the Bureau of International Organization Affairs said they were mistreated and retaliated against by top Trump administration appointees who thought they were -- quote -- "disloyal" to the president. The State Department vowed to provide a corrective action plan within 60 days. In economic news, Wall Street ended this turbulent week of trading on a positive note. The Dow Jones industrial average climbed 306 points to close at 25886. The Nasdaq rose 129 points, and the S&P; 500 added 41.

*(PBS: PBS NewsHour, 2019 (19-08-16),
PBS NewsHour for August 16, 2019)*

One of the clearest functions of the passive voice in the bold-printed sentence above is that it matches the stylistic conventions of general news reporting. As in hard news reporting, this piece prioritizes the event over the agent, and the passive construction is a convenient way to accomplish that. This aligns with what news discourse typically tries to do: provide updates on the status of key developments rather than dwell on who specifically is responsible for each procedural step. By putting the report in the spotlight, the newsreader signals to the viewers that the main thing to pay attention to right now is the existence of this document and the fact that it may soon become public. Hence, if we are to apply Lambrecht's concepts for reading information packaging constructions, the scheme will look like this (2001: 475):

- **Context sentences:** Air Force investigators determined there was insufficient evidence to prove Air Force General John Hyten had a -- quote -- “unprofessional relationship” with his close aide Army Colonel Kathryn Spletstoser. Hyten's polygraph test was also deemed to be inconclusive.
- **Passive voice sentence:** A separate report from the Defense Department's inspector general could be made public as early as next week.
- **Presupposition:** “A separate report from the Defense Department's inspector general could be made x.”
- **Focus:** “public”
- **Assertion:** “x = public”

As per this scheme, the existence of the report itself is pragmatically presupposed. In Lambrecht's terms, this should be labelled as a knowledge presupposition (or K-presupposition) i.e., the reporter assumes that the audience knows, or can easily infer, that the inspector general reports normally exist within such investigations. The speaker's pragmatic assertion “x=public” is what updates and satisfies the hearer's knowledge state. In other words, what is new is not the report, but rather the status of the report. Lambrecht notes that such non-canonical constructions rely heavily on context for interpretability. It would be odd to suddenly mention “a report from the inspector general” if the surrounding discourse had not already prepared the audience to think about official investigations and institutional oversight mechanisms. The surrounding sentences about insufficient evidence, an inconclusive polygraph, and ongoing reviews serve precisely this function: they ensure that the referent “report” is contextually inferable. Lambrecht names this the consciousness presupposition (or C-presupposition) (2001: 475):

An entity or proposition is C-presupposed if the speaker assumes that its mental representation has been activated in the interlocutors' short-term memory at the time of the utterance.

In this passage, the referent "report" is not fully active but is accessible – it definitely corresponds to Prince's categories of "discourse-old" or "inferable" (1992). The audience can easily retrieve the concept of an inspector general report because it fits seamlessly into the bureaucratic frame that the news segment establishes. However, this is not sufficient on its own. For the assertion to work communicatively, the proposition must also be relevant to the current discourse. Here Lambrecht introduces the so-called topicality presupposition (or T-presupposition) (2001: 476):

An entity or proposition is T-presupposed if the speaker assumes that the hearer considers it a center of current interest and thus a likely locus of predication.

The passive construction promotes the report to topic status, making it then the center of predication for the subsequent update ("could be made public next week"). This topicalization is only possible because the C-presupposition has already been satisfied: the audience has the concept of an investigative document active enough to serve as a discourse topic. Hence, the passive sentence builds informational coherence by assuming:

1. that the existence of such a report is known or inferable (K-presupposition),
2. that the report is mentally active in the discourse participants' working memory (C-presupposition), and
3. that it is topical, i.e. a current center of interest (T-presupposition).

Looking closer at the investigation so far, we can see that the triggers within the assertion domain that enable the audience to access the focus denotatum are missing. In other words, it is worth considering how the speaker structures information in such a way that the addressee can recognize, comprehend, or take the intended focus for granted at the moment the utterance occurs. If we isolate a sentence from its surrounding context, and there are no activated referents that cue a presupposed focus, the hearer may struggle to interpret the new information, or may fail to recognize it as an information unit at all. This process depends not only on the presence of inferable or relatable referents, but also on the addressee's prior knowledge of the subject matter. By this, I do not refer to specialized or professional expertise, but rather to general knowledge that is substantial, well-established, and easily retrievable in the hearer's working memory. For

example, if we replace the focus in the passive voice sentence discussed earlier with “*final*”, the listener might experience comprehension difficulty. Questions such as *Does “final” mean the investigation is complete?*, *Does it mean recommendations are being concluded?*, *Does it mean the findings will then be released?*, or *Does it imply prior drafts were incomplete?* would immediately arise, and the intended focus would not be accessible, or at least not fully accessible.

Therefore, the listener’s general, readily accessible knowledge is a key factor in determining whether a particular grammatical construction, especially one designed to foreground information, will successfully convey the intended focus. The speaker must rely on this shared knowledge to ensure that a presupposed focus is interpretable and relevant. Building on Lambrecht’s (2001: 474) concept of unpredictability in focus interpretation, I propose categorizing this presupposed focus into three distinct types, depending on the addressee’s knowledge state at the time of the utterance:

- Fully knowledgeable – the focus component exists in the hearer’s knowledge database and can be evoked immediately at any point.
- Semi-knowledgeable – the focus component exists in the hearer’s knowledge database, but cannot be fully recalled or easily evoked at the time the utterance is produced.
- Unknowledgeable – the focus component is absent from the hearer’s knowledge database and is therefore entirely new information to the addressee.

This distinction emphasizes that the effectiveness of focus-marking constructions depends not only on formal syntax but also on the cognitive and knowledge state of the listener, which guides both interpretation and the pragmatic choice of grammatical structure.

Considering all this, the following elaboration can be suggested (adapted from Lambrecht 2001: 475):

- **Context sentences:** Air Force investigators determined there was insufficient evidence to prove Air Force General John Hyten had a -- quote – “unprofessional relationship” with his close aide Army Colonel Kathryn Spletstoser. Hyten's polygraph test was also deemed to be inconclusive.
- **Passive voice sentence:** A separate report from the Defense Department's inspector general could be made public as early as next week
- **Presupposition:** “A separate report from the Defense Department's inspector general could be made x.”

- K-presupposition: “A separate report from the Defense Department's inspector general could be made x.”
- C-presupposition: “the K-presupposed proposition has been activated”
- T-presupposition: “the K-presupposed proposition is of current interest”
- **Focus:** “public”
- **Assertion:** “x = public”
 - **Presupposed focus:** semi-knowledgeable as the clause suggests that the audience is likely aware that investigations produce formal reports; these exist in the general knowledge database. However, the audience cannot fully recall or evoke the details or timelines of such reports at the moment of utterance. The new element (“as early as next week”) adds specificity that cannot be fully predicted.

The speaker's use of the long passive construction carries several important assumptions. By phrasing the information as “could be made public,” the speaker introduces a sense of possibility, even uncertainty, and leaving room for multiple interpretations. Additionally, the timing, indicated by “as early as next week,” signals that the proposition is current and relevant. This presupposes that listeners are following an ongoing piece of news related to the confirmation process. The presupposed focus relies on the audience's prior knowledge that the inspector general reports exist and play a role in these procedures, though it does not require familiarity with the exact content or schedule of the report. The use of a “separate report” further presupposes that earlier reports have been issued, allowing the listener to situate this update within a broader chain of investigative developments. Still, “separate” means different and this leaves the viewers of the program with a sense of unfamiliarity with the report.

The passive voice itself contributes to a tone of objectivity, which is particularly important when reporting on politically sensitive matters, especially involving allegations. By emphasizing the event, i.e. the report's potential release, rather than the agent responsible, the construction presents the information in a neutral and measured way, consistent with professional journalistic standards. Furthermore, the passive voice here fits into a larger narrative progression. The whole segment consists of short, information-heavy updates from various federal institutions such as the State Department. Several of these updates use agentless or semi-agentless constructions (“was found,” “was deemed,” “were mistreated,” etc.). The passive sentence about

the report seamlessly joins this pattern. It contributes to the segment's stylistic cohesion by matching the informational, non-dramatic, institution-centered tone that runs through the entire piece.

In the following excerpt from COCA, the CBS news report analyzes how Howard Wilkinson, a trader at Danske Bank's Estonian branch, uncovered a massive money-laundering scheme involving around \$230 billion in suspicious transactions. Shell companies like Lantana Trade LLP were carrying out huge trades despite being officially dormant, often routed through offshore jurisdictions such as Russia, the Seychelles, and the Marshall Islands. When Wilkinson raised his concerns, the bank executives took no action, thus prompting his resignation:

It involves nearly a quarter of a trillion dollars of very suspicious money from Russia and the former Soviet Union that was funneled into the Western banking system right under the noses of major banks and regulators in the United States and Europe, who either facilitated it or turned a blind eye. At the heart of it is a whistleblower who found one loose thread and decided to pull on it.

Begin-VT STEVE-KROFT (voiceover): Howard Wilkinson is an Oxford man -- cautious, prudent and a bit of a stickler. **After his cover was blown last fall in a newspaper article as the person who uncovered the scandal, he has spent much of his time wandering the British countryside trying not to be found.**

HOWARD-WILKINSON: Being named as a whistleblower in a case involving dirty Russian money, it's not a good place to be.

STEVE-KROFT: You're still concerned?

HOWARD-WILKINSON: You've got to be, haven't you? The very nature of the people who want to launder money probably means that they're not the sort that you want to go down the pub and have a pint with.

STEVE-KROFT (voiceover): But he did sit down with us and told his tale about a financial crime so big it's hard to fathom.

HOWARD-WILKINSON: The end number that's reported for the whole thing over the six or seven years is two hundred and thirty billion dollars of suspicious money.

(CBS News: 60 Minutes, 2019 (19-05-19), News)

The use of the passive voice in the sentence "After his cover was blown last fall..." serves several important functions in this context. It shifts the focus from the agent who revealed Wilkinson's identity to the event itself (the exposure of his cover). By omitting the agent, the sentence emphasizes the

consequences Wilkinson faced rather than the journalist or newspaper responsible for uncovering this piece of news. This allows the story to center on Wilkinson's vulnerability and the personal risk associated with being a whistleblower. Again, using the passive voice contributes to the tone of objectivity and neutrality. To describe the situation in active voice i.e., "A newspaper blew his cover," would sound more confrontational and even express blame. The passive voice allows the speaker to present the fact of exposure without pointing any fingers, which is particularly important in reporting on sensitive matters such as financial crimes involving large sums of money and international figures. From this, the following scheme of information packaging explanation can be suggested (adapted from Lambrecht 2001: 475):

- **Context sentences:** Howard Wilkinson is an Oxford man -- cautious, prudent and a bit of a stickler.
- **Passive voice sentence:** After his cover was blown last fall in a newspaper article as the person who uncovered the scandal, he has spent much of his time wandering the British countryside trying not to be found.
- **Presupposition:** "After his cover was blown last fall in a newspaper article as x."
 - K-presupposition: "After his cover was blown last fall in a newspaper article as x."
 - C-presupposition: "the K-presupposed proposition has been activated"
 - T-presupposition: "the K-presupposed proposition is of current interest"
- **Focus:** "the person who uncovered the scandal"
- **Assertion:** "x = the person who uncovered the scandal"
 - **Presupposed focus:** fully knowledgeable as the clause suggests that Wilkinson's exposure is salient, comprehensible, and contextually relevant.

The long passive voice sentence above puts in the foreground the event of exposure rather than the actor responsible for revealing Wilkinson's identity. In terms of information packaging, the passive voice allows the narrative to shift attention to Wilkinson and the consequences he faces, rather than the journalist or media outlet. The sentence presupposes that the audience already understands the general context i.e., that Wilkinson is a whistleblower, that he exposed significant financial wrongdoing, and that such exposure carries personal risk.

From a presupposition perspective, the K-presupposition is that the listener knows about the existence of whistleblowers and the risks associated with revealing wrongdoing. The C-presupposition is that the notion of exposure has been activated in the audience's consciousness; earlier context in the segment establishes Wilkinson's role and the ongoing scandal thus making this sentence timely and coherent. The T-presupposition situates the exposure as a center of current interest: it drives the narrative forward and contextualizes Wilkinson's subsequent actions, such as hiding and reflecting on the implications of his whistleblowing.

The presupposed focus of the sentence is the cover being blown, which relies on the audience's prior knowledge of whistleblowing procedures and media reporting. The passive voice allows the focus to remain on the event itself (the revelation) rather than the actor, which is largely irrelevant to the listener's understanding of the consequences. This fully aligns with the journalistic conventions, where objectivity and neutrality are emphasized, especially in politically and financially sensitive stories. The agent is left unspecified, maintaining suspense and probably even tension. Moreover, the choice of the passive construction reflects awareness of the audience's knowledge state. Listeners most likely have general awareness of media involvement and institutional reporting, but not the exact agent responsible for the exposure. By framing the sentence in the passive voice, the speaker ensures that the event can be understood and its implications appreciated without requiring specialized or complete knowledge of the media mechanics involved. Therefore, the fully knowledgeable presupposed focus is effectively activated through contextual cues and prior discourse, which makes the sentence felicitous and informative within the broader narrative.

Finally, in this excerpt from the NPR morning show the anchors are discussing President Donald Trump meeting with the Chinese President Xi Jinping at the G-20 summit in Osaka, who has been aiming to alleviate the trade war situation that had dragged on for about a year until 2019. The U.S. side claimed it was "90 % of the way there" to a deal, though trade experts cautioned that a full agreement was unlikely and instead expected a pause or restart of talks:

WILLEMS: Well, I mentioned before that I think what is going to be determinative here in terms of whether we restart negotiations or whether we have an escalation is going to be how those conditions are presented, how the leaders interact with each other. I think if President Xi comes into the meeting and says absolutely, unequivocally, the U.S. needs to drop tariffs – it needs to fix

Huawei, and it needs to do all these other things – I think that's going to be a bad situation. If, on the other hand, President Xi says, these are my priorities – can you help me out here? Can you show me some flexibility? – I think you could have a productive meeting. And then with Huawei in particular, that is a tricky issue. **And the administration has been clear that Huawei is a national security threat and that action needs to be taken to prevent Huawei from being part of the U.S. network.** Now, whether or not there's flexibility in terms of doing business with the company at all – that may be a different question. I think you theoretically, at least, can distinguish saying, they can't be part of our networks from they can't build cell phones in China, and we're not going to give them any components for their cell phones. Those are different things in my mind.

KING: The Journal also reports that the Chinese will likely ask the U.S. to drop the tariffs it's implemented so far. That likely to happen?

WILLEMS: Not in the short term. What the United States has said – and they've been clear – is that China in the past hasn't followed through on commitments it made.

(NPR: Morning Edition, 2019 (19-06-28), Trump to Talk Trade with China's President at Crucial Weekend Meeting)

The passive voice used in the interview above plays a key role in shaping how the information is perceived. By framing the statement this way, the focus is placed entirely on the necessity of the action itself rather than on who is responsible for carrying it out. The agent (the specific government office or official) is deliberately left unspecified, which allows the sentence to highlight the broader policy imperative. In doing so, the construction draws attention to what must be done rather than who must do it and thus signaling the seriousness of the issue. This use of the passive voice also conveys a sense of objectivity and measured authority. An active phrasing like “the administration needs to take action” would directly assign responsibility and could be read as more politically charged. The passive formulation, in contrast, frames the requirement as a recognized necessity. This gives weight to the institutional consensus rather than the individual initiative. It subtly communicates that the exclusion of Huawei is not merely a personal preference but a matter of formal policy and a systemic concern. From this, the following scheme of information packaging explanation can be suggested (adapted from Lambrecht 2001: 475):

- **Context sentences:** And then with Huawei in particular, that is a tricky issue.
- **Passive voice sentence:** And the administration has been clear that Huawei is a national security threat and that action needs to be taken to prevent Huawei from being part of the U.S. network.
- **Presupposition:** “that action needs to be taken x.”
 - K-presupposition: “that action needs to be taken x.”
 - C-presupposition: “the K-presupposed proposition has been activated”
 - T-presupposition: “the K-presupposed proposition is of current interest”
- **Focus:** “to prevent Huawei from being part of the U.S. network”
- **Assertion:** “x = to prevent Huawei from being part of the U.S. network”
 - **Presupposed focus:**unknowledgeable as the clause suggests that the audience knows generally that policy measures exist for national security but does not have full details of the specific action being discussed.

From an information packaging perspective, the presupposed focus in this sentence is unknowledgeable, because the audience may be aware of Huawei’s controversial status and U.S. concerns about national security, but the specific action that is required i.e., the precise measures to prevent Huawei from participating in U.S. networks, is not something the listener can fully know or recall prior to the utterance. By leaving the focus component unspecified, the speaker introduces new information while relying on the audience’s broader understanding of the context to make sense of it. This statement once again aligns with the journalistic conventions, where the passive is often used to present new or sensitive developments without overburdening the listener with agent-specific details.

The active voice counterpart phrasing such as “the administration needs to take action” would explicitly assign responsibility, which might appear confrontational or partisan related. The passive construction allows the statement to be interpreted as a formal policy necessity and presents the action as a systemic requirement rather than a discretionary choice by any individual. The passive voice version highlights the systemic and procedural nature of the issue. It frames Huawei’s exclusion as part of a larger regulatory and security framework, which is consistent with the audience’s semi-activated awareness of the ongoing U.S.-China trade and technology disputes. By focusing on the action rather than the actor, the sentence guides

the audience to consider the implications of the measure i.e., how it affects network security, trade negotiations, and global technology infrastructure, without being distracted by bureaucratic specifics.

By structuring the sentence this way, the speaker effectively balances the introduction of new information with the audience's prior knowledge, ensuring that the statement is both comprehensible and impactful. The passive voice here is not just a grammatical choice – it is a strategic tool that emphasizes policy necessity, maintains neutrality, and centers the importance of the action in a complex, politically sensitive context.

Limitations

While the present analysis aims to provide a detailed account of how passive voice constructions function as information packaging devices in the spoken discourse, several limitations must be noted. First, the dataset relies on a relatively small, random sample extracted from a single corpus (COCA) which, although large and balanced, offers only mediated transcriptions of speech rather than fully spontaneous conversational data. This limits the degree to which the findings can be generalized to real-time spoken interactions, where hesitations, repairs, and prosody play a crucial role in signaling information status. Second, the study focuses exclusively on passive voice constructions identified via a single query structure (BE+vvN). This excludes short passives, get-passives, and adjectival passives, etc. each of which exhibits distinct discourse functions and may pattern differently with respect to presupposition, topicality, and cognitive accessibility. Finally, the categorization of presupposed focus (fully knowledgeable, semi-knowledgeable, unknowledgeable) is inferential and indirect, given that the COCA corpus lacks metadata on listener background knowledge, attentional state, or conversational goals.

Implications for future research

These limitations nonetheless open avenues for further exploration. Future research would benefit from incorporating multiple corpora to examine how prosody interacts with passive constructions. Expanding the scope to include different passive types, as well as cross-linguistic comparisons, would clarify how structural differences across languages affect speakers' preferred ways of managing information in discourse. This is particularly relevant in light of the findings such as those by Kostadinova (2017), who shows that Bulgarian speakers of English struggle with the passive voice due to structural mismatches and transfer effects. In other

words, the question is whether learners draw on different linguistic signals for marking topicality, focus, or backgrounding (such as word order rather than voice) when producing English discourse. Integrating corpus, experimental, and cross-linguistic perspectives would allow for a more comprehensive theory of how passive voice functions as a cognitive and discourse-pragmatic resource in spoken English.

Conclusion

This study shows that the passive voice plays a much more meaningful role in communication than simply offering an alternative to the active. Based on the analysis, in spoken English, speakers use the passive to keep the topic steady, highlight what is already familiar, and avoid overloading the beginning of the sentence with too much new or complex information. These patterns demonstrate that passive constructions serve as an efficient information packaging tool that aligns the grammatical form with the communicative needs. While the present work is limited to one corpus, it provides a useful starting point for future studies that compare how speakers of different languages make similar information-structuring choices. Understanding these patterns not only deepens our knowledge of how grammar supports communication but can also help improve language teaching.

REFERENCES

- Asami & Sugawara 2023: Asami,D.& Sugawara,S.PROPRES: Investigating the Projectivity of Presupposition with Various Triggers and Environments. In: *Proceedings of the 27th Conference on Computational Natural Language Learning (CoNLL)*,December 6–7, 122–137. Association for Computational Linguistics.
- Banks 2017: Banks, D. The extent to which the passive voice is used in the scientific journal article, 1985-2015. *Functional Linguist*, 4(12). DOI 10.1186/s40554-017-0045-5.
- Birner & Ward 2009: Birner, B. J., & Ward, G. Information structure and syntactic structure. *Linguistics and Language Compass*, 3(4), 1167–1187. <https://doi.org/10.1111/j.1749-818X.2009.00146.x>
- Chankova 2021: Chankova, Y. Information Packaging Effects in Old English Scrambled Double Object Constructions. In: *Studies in the Evolution of the English Language*. Robert Kieltyka (ed.). Berlin: Peter Lang, 2021, pp. 111-128.

- Chankova 2016: Chankova, Y. *Aspects of the Theory of Scrambling*. Blagoevgrad: South-West University Press. ISBN 978-954-00-0088-6.
- Firbas 1992: Firbas, Jan. *Functional Sentence Perspective in Written and Spoken Communication*. Cambridge: Cambridge University Press.
- Hou & Smith 2018: Hou, L. & Smith, D. Modeling the Decline in English Passivization. *Society for Computation in Linguistics*, 1(1), 34-43. doi: <https://doi.org/10.7275/R5ZC812C>
- Huddleston & Pullum 2002: Huddleston, R., & Pullum, G. K. *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press, 2002.
- Kostadinova 2017: Kostadinova, D. *Bulgarian Interferences in English Texts for Specific Purposes*. Blagoevgrad: South-west University Publishing House.
- Lambrecht 1994: Lambrecht, K. *Information Structure and Sentence Form: Topic, Focus, and the Mental Representations of Discourse Referents*. Cambridge: Cambridge University Press.
- Lambrecht 2001: Lambrecht, K. A Framework for the Analysis of Cleft Constructions. *Linguistics*, vol. 39, no. 3, pp. 463–516.
- Mair & Leech 2006: Mair, C. & Leech, G. Current Changes in English Syntax. In A. Aarts & A. McMahon (Eds.), *The Handbook of English Linguistics*, 318–342. Blackwell.
- Prince 1992: Prince, Ellen. The ZPG Letter: Subjects, Definiteness, and Information-status. *Discourse Description: Diverse Linguistic Analyses of a Fund-Raising Text*, ed. by William C. Mann and Sandra A. Thompson, Benjamins, 295–326.
- Roberts 2012: Roberts, C. Information Structure in Discourse: Towards an Integrated Formal Theory of Pragmatics. *Semantics & Pragmatics*, 5(6), 1–69. <http://dx.doi.org/10.3765/sp.5.6>.
- Strunk & White 2002: Strunk, W. Jr. & White, E. B. *The Elements of Style*, 4th edition. Needham Heights, MA: A Pearson Education Company.
- COCA tokens (in the order they appear): PBS 2019: PBS. *PBS NewsHour*: PBS NewsHour for August 16, 2019.
- CBS 2019: CBS. *CBS News: 60 Minutes*. News.
- NPR 2019: NPR. *Trump to Talk Trade with China's President at Crucial Weekend Meeting*. NPR: Morning Edition.

Assist. Prof. Krasimir Spasov, PhD

American University in Bulgaria

Blagoevgrad, Bulgaria

e-mail: kpasov@aubg.edu

<https://orcid.org/0009-0002-0261-6194>