Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A Two-layer Annotation Scheme for the Pragmatic Analysis of the Passive

Emeline Pierre

emeline.pierre@unine.ch

University of Neuchâtel https://orcid.org/0000-0003-0073-9633

Romina Kowsari

University of Neuchâtel

Research Article

Keywords: passive voice, annotation scheme, agent defocusing, information structure, corpus pragmatics

Posted Date: November 4th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11605930/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

A Two-layer Annotation Scheme for the Pragmatic Analysis of the Passive

Pierre, Emeline; Kowsari, Romina

University of Neuchâtel
Avenue du 1er-Mars 26
2000 Neuchâtel Switzerland

Abstract

Current pragmatic research shows that structures such as the periphrastic passive influence the interpretation of utterance meaning, yet none specifies how linguistic structure gives rise to a given pragmatic effect. The existing body of literature remains mainly theoretical and descriptive rather than methodological. This study thus offers a replicable double-layered annotation scheme linking observable linguistic cues such as verb type, lexical polarity, oblique complement, and discursive stakes to interpretive pragmatic categories of agentive attribution and content framing. Inter-annotator agreement reached at least substantial levels across all categories ($\kappa \geq .60$). Applying the framework to a corpus of periphrastic passives from a Corporate Social Responsibility (CSR) report, we show that content framing most frequently foregrounds positive company actions, consistent with the use of the periphrastic passive as a potential resource for managing agent visibility to project a more benevolent image of the company.

Keywords: passive voice, annotation scheme, agent defocusing, information structure, corpus pragmatics

1. Introduction

This paper aims to offer an annotation framework for the pragmatic effects of the periphrastic passive. The construction has long attracted scholarly attention from structural, functional or pragmatic perspectives alike (e.g., Givón 1984; Shibatani 1985, Meluzzi et al. 2021). Each of these approaches seeks to decode the complexities of this structure often reduced to a syntactic transform of its active counterpart or explained solely in terms of patient promotion and agent demotion. Yet the passive is probably best described as multifunctional. Beyond its role in organizing discourse flow and directing the addressee's attention, it is mobilized across various contexts from cooperative vagueness (Mannaioli 2025) in everyday conversation to more strategic forms of concealment in institutional and politically sensitive discourse (Fannes & Claeys 2023, Pierre 2026). Where the current literature on the passive falls short is to specify *why* a given passive contributes to vagueness in a sentence and to deliberate suppression of agency in another, a distinction that the present framework formalizes as part of a broader typology of agentive attribution (Section 4). So far, such a distinction tends to rest on the analyst's intuition and lacks an explicit and replicable procedure, which has direct consequences for how reliably the pragmatic functions of the passive can be studied and compared across corpora.

We propose a two-layer annotation scheme: a first layer establishes seven linguistic features, on the basis of which a second layer annotates the pragmatic effects proper. We test this scheme on a corpus of Corporate Social Responsibility (CSR) reports, a genre in which companies report on their own conduct, under high accountability and reputational pressure, and where the handling of agentive attribution is therefore delicate. By grounding each pragmatic category in observable linguistic criteria and reporting inter-annotator agreement, this scheme offers a replicable alternative to analyst intuition and makes the pragmatic functions of the passive comparable across corpora and studies.

2. Theoretical Background

2.1 From structure to function: the passive beyond syntax.

From a syntactic perspective, the periphrastic passive can be loosely defined as a construction consisting of the auxiliary be and a past participle. Structural accounts generally understand the passive as a process whereby the syntactic configuration of an active clause is altered so that a non-agent is promoted into the

role of 'main topic' of the sentence (Givón 1979:117). In this sense, the passive can be viewed as a formally distinct counterpart of its corresponding active construction.

On of the main functions attributed to the passive in the literature is thus the promotion of the patient or undergoer (e.g., Givón 1979, Shibatani 1988, Gaatone 1999, Carter & McCarthy 2006, Raineri 2010) and the subsequent demotion of the agent, that is usually even supressed from the construction. The agent may be omitted for several reasons (Jespersen 1933): (i) the object is unknown or evident from the context, (ii) there is a communicative motivation for leaving it unexpressed, (iii) the patient is of greater communicative importance.

From a pragmatic perspective, however, the passive cannot be reduced to a syntactic variant of an active construction (Kim 2011:51). Research adopting this perspective has developed along two main lines, one concerned with information flow (Chafe 1976, Gundel 1988), the other with agent-defocusing.

2.2 Information flow and agent-defocusing

In information flow tradition, passive is seen as a grammatical structure that allows controlling information status (i.e. given-new information), topic continuity and end-weight. As Kim (2011: 69) puts it, the correlation between passive constructions and the notions of givenness, newness, and heaviness suggests that passives are strategically exploited by speakers and writers to structure information flow and facilitate the processing and comprehension of the intended message.

A second line of research approaches the passive through the lens of agent-defocusing and agency attribution, especially in sensitive or high-stakes situations, such as those involving blame or responsibility (Meluzzi & et al. 2021). This line of work builds on the evidence that the passive influences readers in ways that extend beyond the literal content of the message (Chan & Maglio 2020). One account of this effect draws on construal level theory, according to which psychologically distant events are represented in more abstract terms and close events in more concrete terms (Trope & Liberman 2010). As passive constructions foreground the patient while backgrounding the agent, they have been argued to lead readers to perceive the reported event as more distant, which in turn activates a higher, more abstract level of construal compared to the same informational content in the active voice (Trope & Liberman 2000 as discussed in Chan & Maglio 2020). This shift in construal level has consequences that extend beyond the immediate

interpretation of the utterance and affect decisions and judgements in domains such as the attribution of blame and responsibility.

Meluzzi et al. 2021 provide experimental evidence that voice alternation can affect blame attribution. In reports of gender-based violence, victims are assigned greater blame when the incident is described in the passive voice. Bohner (2001) reports a comparable, though more conditional, effect. Passive descriptions of a rape incident increased negative perception of the victim among readers who endorsed rape myths (i.e. general stereotypical beliefs that minimize offender guilt and hold victims responsible (Burt, 1980)), while the same descriptions elicited greater empathy and more positive evaluations of the victim among readers who did not. A related pattern has been documented in institutional settings by O'Nyangeri et al. (2024). The authors show the impact of grammatical structures in courtroom interpretation and translation, with direct consequences for how testimony is understood by the court. Taken together, these findings suggest that the choice between active and passive constructions bears on how hearers infer who is responsible for the reported event. These findings, however, come from experimental settings, in which the effect of voice is measured on an audience. The present framework transposes this insight to a corpus-annotation setting, where the pragmatic effect assigned to an occurrence results from a criteria-based inference.

2.3 Towards a typology

What remains open, across this body of work, is how one determines, for a given occurrence, which pragmatic effect is triggered by the use of the periphrastic passive. Existing typologies address this question only partially. Shibatani (1985) and Myhill (1997) distinguish structural subtypes of the passive and, in Myhill's work (1997), several categories of agent. But the pragmatic effect associated with each subtype is not derived from criteria that could be applied systematically. Lingle (2021) tests an analytical framework based on reader response to identify the conditions under which an omitted agent remains unrecoverable to the reader, but this proposal only targets a single effect, mystification, and does not apply to other effects. To the best of our knowledge, no existing proposal, thus, specifies the linguistic criteria by which a passive could be assigned to one pragmatic effect rather than another, a gap the annotation scheme developed in the Section 4 addresses.

3. Data and Corpus

The data for this study consist of the Corporate Social Responsibility (CSR) report of Inditex, a Spanish multinational clothing company. It was published in 2024, in English, and is freely available on the internet. The CSR report was selected as the target genre as it constitutes a paradigmatic case of institutional disclosure text where accountability and reputational pressures make the passive a likely rhetorical resource. The report has been processed in full, from which 453 instances of the periphrastic passives were extracted by spaCy, an open-source library for Natural Language Processing in Python. All extracted instances were manually checked to exclude adjectival passives and false positives.

4. Proposed Annotation Scheme

4.1. Rationale

Research instruments matter because, in their absence, pragmatic studies tend to vary uncontrollably across analysts and corpora (Kasper and Dahl 1991) and the annotation of pragmatic phenomena already requires a great deal of interpretation on the part of the analyst (De Felice & Strik-Lievers 2024). We, therefore, propose a two-layer annotation scheme to annotate the use of the passive. The first layer includes the diagnostic criteria, i.e. text-based features in the span[1] of the passive, annotated independently. Layer 2 consists of the pragmatic effects produced by the use of the passive, based on the first layer of annotation. The value(s) taken by the features in Layer 1 condition the interpretation of one pragmatic effect rather than another in Layer 2.

4.2. Layer 1 — Linguistic Categories

The present section offers a description of the diagnostic criteria of the first layer. These criteria are text-based and defined so as to require as little inference as possible. Agent_status, Verb_type and Numerical_statistical_content rest mostly on formal or lexical identification, whereas Lexical_polarity, Oblique_compelment and Discursive_stakes call for a limited judgement, constrained by the decision rules given below. After a corpus-driven process and an initial phase of manual annotation on a subset of 10% the corpus, we identified seven linguistic features relevant to the annotation of the passive and its pragmatic effects. These features are all annotated independently. Layer 1 values do not mechanically

[1] The span covers the sentence where the auxiliary *be* + participle is located.

determine Layer 2 categories. They define typical profiles that guide the analyst. Where a profile is compatible with more than one category, the priority notes given in Section 4.3 apply.

4.2.1 Agent_status

Agent is coded with a binary value: *explicit* or *absent.* The value *explicit* applies when the agent is syntactically realized through the prototypical *by*-phrase complement, such as in 'The plan was designed by the company'. The value *absent* applies when the agent is not syntactically realized (either recoverable from the immediate co-text or not), as in the following examples:

- *Mistakes were made:* agent absent, not recoverable from the co-text.
- *Mistakes were made in our value chains;* agent absent (still no *by*-phrase), though partly recoverable.

Both instances are coded as *absent* under this criterion, regardless of the agent's recoverability.

4.2.2 Verb_type

The classification of the verb selected within the passive construction adds a semantic dimension to the framework: certain verb types recur with certain pragmatic effects, a pattern that motivated the inclusion of this criterion. Table 1 presents each of the six values distinguished, with its definition, and representative examples.

Table 1 Values of verb_type

Value	Definition	Example
Enunciative	Speech/assertion verb reporting stance	*confirm, state, indicate (without textual locative complement), note, announce*
Positive_achievement	Achievement with positive valence	*agree, grant, implement, launch, award*
Formulaic_editorial	Self-referential to the document; locates information within the text	*provide, present, show, include, indicate (typically followed by a textual locative complement)*
Procedural_reporting	Operational process internal to the company,	*send, reuse, recycle*
Impact_causation	Causation of an effect or impact	*produce, cause, affect, generate*
Other	Verbs not covered by the above categories	—

Note: disambiguation for the *formulaic_editorial* relies on the presence of a textual locative complement (e.g., *below*, *above*, *in this report*), not on the verb alone. A verb from the list is coded as *formulaic_editorial* when it is followed by such a complement, referring to a location within the document itself.

- *An estimate is provided below the table*; locative complement (*below the table*), coded as *formulaic_editorial*
- *Training is provided to all new employees*; complement denotes a recipient (*to all new employees*), not a textual location, not coded as *formulaic_editorial*

4.2.3 Lexical_polarity

Lexical polarity is coded according to the presence and orientation of evaluative lexis (Martin & White 2005) (noun, noun phrase, adjective and/or adverb) in the immediate context[2] of the passive (Wilson et al.

[2] Unless specified, the immediate context refers to the span itself. Lexical_polarity may draw on the preceding sentence in cases where the evaluative lexis referring to the passive is not contained in the span itself.

2009) indicating the semantic orientation of the content. Semantic orientation is treated as bi-directional (positive/negative) and admits degrees of intensity (mild to strong) (Turney & Littman 2003), although intensity is not coded in this project. Three values are coded: absent/positive/negative

Negative: Detection of terms with negative semantic orientation (e.g., extreme weather, severity of each event)

- *In order to weigh the financial impacts of extreme weather events caused by climate change, the changing probability and severity of each event is used to quantify the increase or reduction of the physical impacts expected at facility level.*

Positive: Detection of terms with positive semantic orientation (e.g., new well-being committees)

- *During 2024, new Well-being Committees were established in markets such as France, Serbia, Switzerland and United States.*

4.2.4 Quantifier_scope

This criterion codes the presence of a quantifier or quantificational expression (see Szabolcsi 2010 on quantification). Following Keenan & Stavi (1986), quantifiers are treated as scalar (Horn 1989), starting from the lower end of the scale or the absence of an entity or phenomenon to the upper end of the scale, that is, extending to the whole situation. For the purpose of this study, the criterion includes three categories:

Table 2 Value of Quantifier_scope

Value	**Definition**	**Example**
Absent	No quantifier or quantificational expression	_
Totalizing	Quantifier that occupies the upper end of the scale and extends the scope of action	*Entire, all, fully*
Negative	Quantifier that occupies the lower (null) end of the scale, typically bearing on the absence of wrongdoing or shortcoming.	*No, none, zero, 'no fine was imposed'*

Mid-scale quantifiers (e.g., some, several, most) are not coded, as only the endpoints support the claim-legitimation effect (Section 4.3.2.1).

4.2.5 Oblique_complement

Delbecque (2014:219) sees prepositions as devices for agenthood management and suggests that prepositional complements other than the prototypical *by*-phrase can be used to express agentivity. The criterion *Oblique_complement* codes the presence of a non-prototypical agentive participant in the immediate context of the passive, that is, a prepositional complement that points to a causal factor associated with the event. Unlike the prototypical *by*-phrase, which encodes the agent proper (see *Agent_status* criterion), these obliques provide agentivity-related information (actors, sources, locations, or instruments), without receiving the full agentive status.

Table 3 Values of Oblique_complement

Value	**Definition**	**Example**
Absent	No oblique complement present	_
Present	Causally relevant participant realized by a non-prototypical agentive complement.	*The decision was taken in Brussels. / This is obtained from animals. / Peace will not be obtained with weapons.*

Agentive complements introduced by *by* are excluded from this criterion and coded exclusively under *Agent_status*, since Delbecque's (2014) account reports the *by*-phrase as distinct from other obliques in its capacity to encode the agent proper. A complement counts as *Present* only when it supplies, directly or by metonymy, a candidate to "who or what is behind the event?". Temporal adjuncts are excluded.

4.2.6 Numerical_statistical_content

Presence of numerical content in the immediate context of the passive. This content provides information on the substance of the action described (procedures, quantities, achievements or details). This excludes temporal refernces such as calendar dates or years.

The value is coded : yes/no

- *1,892 MWh of biomethane were generated : yes*
- *The report was published in 2024 : no*

4.2.7 Discursive stakes

Discursive_stakes codes whether the passive occurs in topics likely to call for mitigated formulation, that is, topics in which speakers are likely to reduce their exposure or commitment in response to face-threatening or sensitive content (Fraser, 1980; Caffi 1999). Two values are distinguished: *high* and *low*. The value *high* applies when the passive occurs within one of the following stake-sensitive domains: human rights, discrimination, worker safety and health, vulnerable population, environmental harm and forced labor. This list was established during the pilot phase and is specific to the present corpus. Other genres may call for a different list. The value *low* applies to all other domains.

- *In this way we aim to listen to those people who are least represented and who may* ***be most affected****.*

This criterion is evaluated over a three-sentence window (preceding sentence – span -following sentence), given that topic-level features are less reliably identifiable in the span alone.

4.3. Layer 2 — Pragmatic Categories

Layer 2 includes two independent features, which are not necessarily mutually exclusive, and which build on the first layer of annotation. The corpus-driven exploratory analysis revealed that these two features, or dimensions, answer different questions. The first concerns agentive attribution i.e., who the agent is, how responsibility is treated, while the second concerns how the propositional content is framed to guide the audience's interpretation. The two dimensions may, but need not, co-occur in a single occurrence of the passive as will be illustrated below.

The first dimension (Agentive_attribution) includes the categories *unmarked, vagueness, agent-defocusing, shifting responsibility* and *distancing.* The second dimension (Content_framing) includes *claim_legitimation, focusing on positive actions, focusing on details*.

4.3.1 Features related to Agentive_attribution

4.3.1.1 *Unmarked*

The category *unmarked* refers to cases in which the passive is used to meet genre-specific stylistic expectations such as the impersonal register conventionally associated with formal report writing, such as CSR reports. In occurrences classified as *unmarked*, the passive fulfills discourse-organizational functions and indicates where information is located in the report, for instance by pointing the reader to a talbe or a section. In these cases, the agent is not relevant to the communicative purpose of the utterance, and its omission follows from Grice's (1975) maxim of quantity and the relevance-theoretic principle (Sperber & Wilson 1986) of optimal relevance: no more information than the genre requires is supplied.

- "An estimate of this waste generation **is provided** below the table"

Typical Layer 1 profile for *unmarked:*

Agent_status : absent

Verb_type: formulaic_editorial

Discursive_stakes: low

This profile can be contrasted with the same verb in a context where the passive is not used to organize information in the report as in *Training is provided to all new employees*. This occurrence of the passive will not be coded as *unmarked*, since *is provided* here does not serve to situate information in the report and does not co-occur with a textual locative complement (see Verb_type, Section 3.2.2).

4.3.1.2 *Vagueness*

Our definition of the criterion vagueness as a pragmatic effect produced by the passive draws on Mannaioli (2025) who treats vagueness as a discursive device through wich some content is left unsaid (2025:10), intentionally or not, since the speaker's intentions are not directly accessible to the analyst. More precisely, the periphrastic passive can be analysed as an instance of syntactic vagueness defined as the omission of a syntactic constituent whose content is discursively relevant but not retrievable from the context (Mannaioli 2025: 86; Cominetti et al. 2023).

Vagueness can be cooperative and non-cooperative. In some contexts, it facilitates discourse management by allowing the speaker to remain appropriately imprecise where precision is not yet warranted (Mannaioli

2025:60). This use of the periphrastic passive is different from *unmarked* as it does not refer to discourse-organizational functions and does not orient the reader within the document itself. Put simply, cooperative *vagueness* refers to the report of situations for which naming the agent would not add discursively relevant information in the context and for which nothing is at stake. The non-cooperative use corresponds to our category *agent-defocusing* and will be discussed in Section 4.3.1.3.

Typical Layer 1 profile for *vagueness* :

Agent_status: absent

Verb_type: any value other than formulaic_editorial

Discursive_stakes: low

- "Each brand independently sorts the reasons for customer contacts, according to their specific needs, although these reasons **can be grouped** into four broad areas"

4.3.1.3 *Agent-defocusing*

In non-cooperative contexts, the same mechanism can serve to hide (the lack of) information (Kaufer 1983), or to encode content only implicitly (Lombardi Vallauri 2019). As a result, the reader is left in a state of "unstructured indecision about what remains unsaid" (Kaufer 1983: 211). When the passive suppresses an agent, whose identity is discursively relevant, underspecification stops to be a neutral packaging choice and may serve as a strategic device for managing what the reader can access.

What distinguishes *agent-defocusing* from vagueness is the stakes related to the domain in which the omission occurs. The passive will receive a cooperative interpretation in low-stakes contexts (*vagueness*) and a strategic one in high-stakes contexts (*agent-defocusing*). In high-stakes contexts, a further distinction applies. *Agent-defocusing* covers cases in which the agent is omitted with no alternative participant proposed in its place through an oblique complement. The latter case will be treated separately under *shifting_responsibility* (Section 4.3.1.4). The example below illustrates the former case. It addresses the topic of minority representation ('people the least represented') that is, a potentially high-stakes, sensitive context, where the identity of the agent affecting the people described is discursively relevant. Readers arguably need this information to evaluate the claim.

- "In this way we aim to listen to those people who are least represented and who **may be most affected**."

Typical Layer 1 profile for *agent-defocusing:*

Agent_status : absent

Discursive_stakes : high

Oblique_complement : absent

4.3.1.4 *Shifting responsibility*

The category *shifting_responsibility* is similar to *agent-defocusing*, in that the agent remains absent from the periphrastic passive but it differs in that the agentive attribution or responsibility is explicitly redirected toward another participant through an oblique complement (Section 3.2.5). Following Delbecque's (2014) account of oblique complements as devices of agenthood management, this redirection gives the designated participant a degree of causal relevance without assigning it full agent status.

- "Mistakes **were made** in our value chains"

The oblique complement *in our value chains* directs the reader toward a candidate agent. In this case, responsibility is not totally omitted, as in *agent-defocusing*, but displaced.

Typical Layer 1 profile for *shifting_responsibility:*

Agent_status : *absent*

Discursive_stakes : *high*

Oblique_complement: *present*

4.3.1.5 Distancing

In distancing, the agent may be recoverable or self-evident (here, the speaker or the company itself), and the periphrastic passive produces an effect of depersonalization. The speaker presents themselves as a neutral relay of information (Goffman 1981). The rhetorical effect may include, for instance objectivity or institutional authority. The action appears as an automatic process. Vagueness creates an information gap while distancing constructs an enunciative stance.

- "In this context**, it is confirmed** that for now the Group's main activity has stopped"

The addressee can easily infer that it is the Group itself that confirms.

Typical Layer 1 profile for *distancing* :

Agent_status : *absent*

Verb_type: *enunciative* ; rephrasable with *we*

4.3.2 Features related to Content_framing

4.3.2.1 Claim_legitimation

The use of the passive combined with extreme case formulations (ECF) (Pomerantz 1986), or expressions occupying the endpoints of a quantificational scale (Horn 1989) contribute to legitimizing a claim by making it more resistant to challenge or refutation.

Two subtypes are distinguished, which correspond to the two poles coded under Quantifier_scope (Section 4.2.4).

The first subtype, *claim_legitimation_negation*, involves the negative pole of the scale and foregrounds the absence of a negative outcome. The following example illustrates this pattern.

- “This year, no fines, penalties, or damages were imposed for discrimination or human rights infringements materializing in 2024 in connection with our people."

This utterance foregrounds the absence of any sanction and implies that the company has had no adverse impact on human rights. However, the absence of penalties does not entail the absence of infringements.

The second subtype, *claim_legitimation_totalizing*, involves the upper end of the scale with totalizing quantifiers. Consider the following example:

- “"Furthermore, the scope of control **is extended** to cover the entire inventory of chemicals for manufacturing (dyes, pigments, and ancillary chemicals used in textile manufacturing), except for commodities, cleaning products, and products used to purify water."

The periphrastic passive removes any identifiable agent and is present as an objective fact.

Simultaneously, the quantifier *entire* amplifies the scope of the claim and suggests total coverage. Not

some chemicals but the entire inventory. The combination of the two mechanisms results in a claim that is not easily attributable nor falsifiable.

Typical Layer 1 profile:

Agent_status : *Absent*

Claim_legitimation_negation : Quantifier_scope: *negative*

Claim_legitimation_totalizing : Quantifier_scope*: totalizing*

4.3.2.2 Focusing on positive action

The periphrastic passive allows the foregrounding of positive actions, whereby readers first access the situation through positive information with the subsequent backgrounding or full omission of the agent. This process allows the brand to give more weight to its favorable conduct and contributes to the construction of a positive institutional ethos. This pattern echoes Coppola & Greco (2024:214) analysis of Italian banking communication, where the institution portrays itself as an agent that creates things and supports (people and initiatives, or itself) by appealing to competence and benevolence.

Example from our corpus: “Additionally, in France, a *new seniority bonus* for 25 years with the Company **has been agreed** with the representative trade unions.”

Typical Layer 1 profile:

Verb_type: *positive_achievement* and/or Lexical_polarity: *positive*

4.3.2.3 Focusing on details/numbers

This use of the periphrastic passive foregrounds procedural detail and numbers or statistics. The effect is that readers focus on the foregrounded figure and not on the recovering of the agent. Consider the following example:

- "8,477 people were trained in work teams”

The foregrounding of a statistic (8,477 in the example) directs attention toward the number itself. Who trained them or how is left unaddressed. Such details carry an implicit expectation of relevance (Grice 1975), suggesting accountability and engagement. But this salience works against the retrieval of the agent as the reader's attention is held by the figure and the defocused agent behind it receives less attention.

Typical Layer 1 profile :

Numeric_statistical_content : yes

Verb_type : procedural_reporting

4.3.3 Decision rules

Several Layer 1 profiles are compatible with more than one category. Categories are therefore assigned by applying the rules below in the order given. The first rule that applies determines the category.

Agentive_attribution concerns the absence of the agent, so periphrastic passives with an explicit agent (Agent_status: explicit) receive no label. For the remaining occurrences, unmarked applies when Verb_type is formulaic_editorial and Discursive_stakes is low. If not, distancing applies when Verb_type is enunciative and the passive can be rephrased with we as agent without changing the content (it is confirmed can be rephrased as we confirm). If not, and Discursive_stakes is high, shifting_responsibility applies when an oblique complement is present, and agent-defocusing when none is present. All other occurrences, which have low stakes, are coded vagueness.

For Content_framing, focusing on positive action applies when Verb_type is positive_achievement and/or Lexical_polarity is positive. If not, claim_legitimation applies when Quantifier_scope is totalizing or negative. If not, focusing on details applies when Numerical_statistical_content is yes. Otherwise, no label is assigned. This order gives priority first to the framing carried by the lexis of the whole utterance, then to the framing that depends on a specific lexical trigger, and finally to the framing that is compatible with the other two.

5. Reliability and Validation

Since studying the pragmatic effects of a phenomenon relies heavily on human interpretation, the coding procedure must be as consistent as possible across analysts. Reliability was assessed in two phases. In Phase 1, a random selection including 10% of the instances was independently annotated by two coders using the annotation framework proposed in Section 4. The second coder, a master's student, was trained beforehand on a separate set of data, not included in the test sample to ensure a shared understanding of the scheme and all possible difficulties or underspecification were discussed and clarified in the scheme. Layer 2 labels were assigned on the basis of each coder's own Layer 1 annotation. Cohen's kappa (Cohen 1960) was calculated to assess the degree of agreement between the coders and hence the reliability of the annotation framework (see Table 4).

Category	**Phase 1 Cohen's κ**	**Phase 1 % agreement**	**Phase 2 Cohen's κ**	**Phase 2 % agreement**
Agent_status	0.78	95.7%	-	-
Verb_type	0.62	72.3%	0.79	84.2%
Lexical_polarity	0.46	76.6%	0.61	81.6%
Quantifier_scope	0.76	93.6%	-	-
Oblique_complement	0.78	89.4%	-	-
Numerical_statistical_content	0.83	95.7%	-	-
Discursive_stakes	0.62	87.2%	0.65	89.5%
Agentive_attribution	0.54	72.3%	0.70	81.6%
Content_framing	0.69	78.7%	0.84	92.1%

Table 4. Inter-rater agreement results

Categories with κ below .70 in Phase 1 were revised thorough the clarifications and decision rules given in Section 4 and tested again in Phase 2 on a new random sample. The other categories were not re-tested (indicated by "-"in Table 4). By the end of the annotation process, all categories reach at least substantial agreement ($\kappa \geq .60$), which is considered an adequate threshold for reliable manual annotation in corpus pragmatics (Landis & Koch, 1977) and lends confidence to the analyses reported in the remainder of this article.

6. Illustrative analysis

This section presents an illustrative case of how the proposed framework applies to the analysis of periphrastic passives and information backgrounding in CSR reports. We zoom in on two parameters *focusing on positive actions* (Layer 2) and *discursive_stakes* (Layer 1) and then discuss their interpretive implications for the corpus under study.

Among the three values of *content_framing*, focusing *on positive actions* is the most frequent in the annotated corpus, occurring in 90 of the 453 periphrastic passive occurrences. This frequency is worth examining in detail, as positive action framing shows one of the most visible interactions between the two annotation layers.

Layer 1 annotation identified, within the span, verb types coded as *positive_achievement* alongside nouns, noun phrases, or adjectives whose polarity was positive. The co-occurrence of these elements with a passive form contributes to creating a benevolent image of the company which shifts readers' attention toward accomplishments and favorable conduct in broader terms, as illustrated in the following Example (1) :

(1) “Moreover, to improve work-life balance, additional rest measures **were approved** for full-time and part-time employees in the United States."

The verb "were approved" is coded as *positive_achievement*, which places the approval itself in the foreground. The approving body, presumably the company’s management, remains unnamed. The passive makes the measure, and not its author, the topic of the utterance. The company’s caring stance is thus foregrounded, while the omission of approving agent backgrounds institutional responsibility, defocuses the role of management and directs attention instead to the positive achievement itself.

A similar pattern appears in Example (2):

(2) “(...) we nurture inclusive environments where everyone feels welcome and where prejudice, gender stereotypes and discrimination based on gender identity and sexual orientation **are addressed**."

The verb phrases *nurture inclusive environments* and *feels welcome* foregrounded in the span, increase the polarity of the utterance toward a positive connotation and creates a positive image in readers' interpretation.

The passive *are addressed* leaves unspecified who is responsible for addressing prejudice, gender stereotypes and discrimination.

Content coded with *positive_achievement* as *Verb_type* and/or *positive lexical_polarity* Layer 1 motivates the classification of the occurrence as focusing on positive actions in the *Content_framing* feature in Layer 2. As the topic is discrimination, *discursive_stakes* is *high*, and with no oblique complement, *agentive_attribution* is *agent-defocusing*.

A second criteria that is especially illustrative of the importance of the bridge between Layer 1 and Layer 2 is *discursive_stakes*. This cue conditions the selection of the subcategory within *Agentive_attribution in Layer 2.* In Layer 1, *discursive_stakes* is coded as *high* when the span concerns a sensitive domain including topics such as human rights, discrimination, workers' safety and health, vulnerable populations, or environmental harm. Where such stakes are high, the agent's identity is more likely to be mitigated for question of responsibility, blame avoidance or plausible deniability. Accordingly, discursive stakes influences the second layer's annotation process in the agentive_attribution feature, in that an utterance that includes sensitive content would be categorized as *agent-defocusing* in the case of no oblique complement or as *shifting_responsibility* if another agent is present or when an oblique complement redirects responsibility toward another participant in the utterance, whether or not that participant is itself agentive. Consider Example (3).

(3) "In 2024, no inherently severe incidences of human rights violations in our supply chain have been detected through this channel,"

The expressions of *severe incidences* and *human rights violations* both increase the sensitivity of the content, coded as high under *discursive_stakes* in Layer 1. For the second layer, as there is an oblique complement, "through this channel," the responsibility of dealing with this sensitive content shifts from the company itself to a detection mechanism, which in this case would be "this channel", rather than a participant accountable for the reported events. This displaces the locus of responsibility onto procedural or instrumental detail, without assigning it to an accountable participant. The same occurrence is coded *claim_legitimation* for *Content_framing* as the negative quantifier *no* foregrounds the absence of incidents, although the absence of detection through one channel does not entail the absence of violations.

Conversely, the absence of high discursive stakes can lead to an occurrence to be coded as *vagueness* or *unmarked* in Layer 2. A f*ormulaic_editorial verb* points to the stylistic conventions of formal and legal reports *(unmarked)* which has been used in 44 occurrences out of the total number of 453; in the absence of both high *discursive_stakes* and *formulaic_editorial* verb, the utterance is more likely to reflect a case where naming the agent would not add discursively relevant information *(vagueness)*.

Taken together, these two cases illustrate how Layer 1 values orient the assignment of Layer 2 categories, and how the two Layer 2 dimensions can combine in a single occurrence.

7. Periphrastic passives and CSR reports

The framework was applied to the Inditex CSR to examine potential strategic communication patterns, particularly regarding how responsible agents are defocused and responsibility is shifted. CSR reports are public disclosures companies publish about their environmental, social, and governance performance. Prior research shows that their language is often strategically crafted and not purely descriptive. Kochkina et al. (2024) found that firms tailor sector-specific vocabulary, such as "compliance" in banking or "emissions" in energy, to align sustainability narratives with stakeholder expectations. Similarly, Nwagbara and Belal (2019) showed that companies use inclusive pronouns, positive framing, and persuasive phrasing to construct an image of a "responsible organisation," sometimes masking gaps between rhetoric and actual practice. Both studies link this strategic language use to impression management and the risk of greenwashing (Kochkina et al., 2024; Nwagbara & Belal, 2019). The observations of the present analysis are consistent with this literature while pointing to a linguistic mechanism largely unaddressed in it: the use of the periphrastic passive to foreground favorable outcomes and manage the visibility of responsible agents in sensitive domains, for the company’s benefit.

8. Conclusion

This methodological paper offers a novel double-layer annotation framework to analyze periphrastic passives in corporate discourse. It shows how the linguistic elements included in their span can lead to different pragmatic effects in the audience. To the best of our knowledge, a comprehensive annotation

framework capable of analyzing periphrastic passives, specifically by bridging syntactic-semantic analysis with pragmatic effects, remains absent from the literature, which points to a research gap in pragmatic research. Through this double layer annotation, supported by an interrater agreement, our framework aims to provide a more transparent and replicable analysis of periphrastic passives and shows how specific word choices orient the pragmatic effects that a passive makes available to readers. Layer 1 codes seven text-based criteria within the span of the passive: Agent_status, Verb_type, Lexical_polarity, Quantifier_scope, Oblique_complement, Numerical_statistical_content and Discursive_stakes. Layer 2 uses these values and the decision rules of Section 4.3.3 to code two interpretive dimensions, Agentive_attribution and Content_framing. Applied to the 453 periphrastic passives of Inditex's 2024 CSR report, the scheme reached substantial to almost perfect agreement on both Layer 2 dimensions in the second phase ($\kappa = .70$ and .84), and the illustrative analysis showed that the two dimensions can combine in a single occurrence, as in Example (2), where focusing on positive action co-occurs with agent defocusing.

This framework provides a novel approach to pragmatic effects, as it shows that syntax, semantics, and utterance framing produce different effect that can manipulate (in the technical meaning of the term) readers' attention and interpretation and can therefore potentially leads to a strategical use. The second layer relies on a more interpretive evaluation of the pragmatic effect, which makes assumptions about how readers' attention and interpretation can be affected by the syntactic structure of the utterances in the first layer. Although these assumptions are based on the analyses of the first layer, the subjectivity of the analysis in the second layer might count as a limitation of this methodology. As it is impossible to investigate the actual pragmatic effect and the interpretation in the reader's mind, or to access communicators' intentions, the assumptions about strategic communication and intentional content framing should be taken with caution. The corpus consists of one report from one company, and some criteria, such as the list of high-stakes domains and the coding of numerical content, were developed for CSR reporting and may need adjustment elsewhere. For future research we suggest applying this framework to other genres with strong conventions of impersonality, such as legal, scientific, and institutional communication. Reception studies, building on experimental work on the effects of voice (Meluzzi et al. 2021; O'Nyangeri et al. 2024), can complement this framework and test whether the Layer 2 categories predict how readers interpret the passive.

References

Bohner, G. (2001). Writing about rape: Use of the passive voice and other distancing text features as an expression of perceived responsibility of the victim. *British Journal of Social Psychology, 40*(4), 515–529.

Burt, M. R. (1980). Cultural myths and supports for rape. *Journal of Personality and Social Psychology, 38*(2), 217–230.

Caffi, C. (1999). On mitigation. *Journal of Pragmatics, 31*(7), 881–909.

Carter, R., & McCarthy, M. (2006). *Cambridge grammar of English: A comprehensive guide; spoken and written English grammar and usage*. Cambridge University Press.

Chafe, W. L. (1976). Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In C. N. Li (Ed.), *Subject and topic* (pp. 25–55). Academic Press.

Chan, E. Y., & Maglio, S. J. (2020). The voice of cognition: Active and passive voice influence distance and construal. *Personality and Social Psychology Bulletin, 46*(4), 547–558.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*(1), 37–46.

Cominetti, F., Gregori, L., Lombardi Vallauri, E., & Panunzi, A. (2024). IMPAQTS: A multimodal corpus of parliamentary and other political speeches in Italy (1946–2023), annotated with implicit strategies. In *Proceedings of the IV Workshop on Creating, Analysing, and Increasing Accessibility of Parliamentary Corpora (ParlaCLARIN) @ LREC-COLING 2024* (pp. 101–109).

Coppola, C., & Greco, S. (2024). Argumentation and agency for building ethos in an Italian bank's institutional website. In *Proceedings of the Tenth Conference of the International Society for the Study of Argumentation* (pp. 205–217). Sic Sat.

De Felice, I., & Lievers, F. S. (2024). Building a pragmatically annotated diachronic corpus: The DIADIta project. In *Proceedings of the Tenth Italian Conference on Computational Linguistics (CLiC-it 2024)* (pp. 266–272).

Delbecque, N. (2014). Passive voice and causal roles in Spanish. In N. Delbecque, K. Lahousse, & W. Van Langendonck (Eds.), *Non-nuclear cases*. John Benjamins.

Fannes, G., & Claeys, A.-S. (2023). Shaping attributions of crisis responsibility in the case of an accusation: The role of active and passive voice in crisis response strategies. *Journal of Language and Social Psychology, 42*(1), 3–30.

Fraser, B. (1980). Conversational mitigation. *Journal of Pragmatics, 4*(4), 341–350.

Gaatone, D. (1999). Pour une définition restrictive du passif en français. In *La langue, les signes et les êtres: Actes du colloque de l'Institut d'Études Romanes de l'Université de Copenhague, le 3 octobre 1998. Le passif* (Études romanes, 45, pp. 15–22). Museum Tusculanum Press.

Givón, T. (1984). Prolegomena to discourse-pragmatics. *Journal of Pragmatics, 8*(4), 489–516.

Goffman, E. (1981). *Forms of talk*. University of Pennsylvania Press.

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

Gundel, J. K. (1988). Universals of topic-comment structure. In M. Hammond, E. A. Moravcsik, & J. Wirth (Eds.), *Studies in syntactic typology* (pp. 209–239). John Benjamins.

Horn, L. R. (1989). *A natural history of negation*. University of Chicago Press.

Jespersen, O. (2003). *Essentials of English grammar* (25th impression). Routledge. (Original work published 1933)

Kasper, G., & Dahl, M. (1991). Research methods in interlanguage pragmatics. *Studies in Second Language Acquisition, 13*(2), 215–247.

Kaufer, D. (1983). Metaphor and its ties to ambiguity and vagueness. *Rhetoric Society Quarterly, 13*(3–4), 209–220.

Keenan, E. L., & Stavi, J. (1986). A semantic characterization of natural language determiners. *Linguistics and Philosophy, 9*(3), 253–326.

Kim, K. (2011). A pragmatic-functional account for passive constructions in discourse. *The Jungang Journal of English Language and Literature, 53*(4), 51–72.

Kochkina, N., Macchia, S., & Floris, M. (2024). Strategic language use in sustainability reporting: An empirical study. *Sustainability, 16*(23).

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159–174.

Lingle, W. (2021). When do agentless passives mystify social actors in the minds of readers? *Critical Discourse Studies, 20*(2), 150–165.

Lombardi Vallauri, E. (2019). Implicitation and power of choice. *Rivista italiana di filosofia del linguaggio, 13*(2).

Mannaioli, G. (2025). *Vagueness as an implicitating persuasive strategy*. John Benjamins.

Martin, J. R., & White, P. R. R. (2005). *The language of evaluation: Appraisal in English*. Palgrave Macmillan.

Meluzzi, C., Pinelli, E., Valvason, E., & Zanchi, C. (2021). Responsibility attribution in gender-based domestic violence: A study bridging corpus-assisted discourse analysis and readers' perception. *Journal of Pragmatics, 185*, 73–92.

Myhill, J. (1997). Toward a functional typology of agent defocusing. *Linguistics, 35*(5), 799–844.

Nwagbara, U., & Belal, A. (2019). Persuasive language of responsible organisation? A critical discourse analysis of corporate social responsibility (CSR) reports of Nigerian oil companies. *Accounting, Auditing & Accountability Journal, 32*(8).

O'Nyangeri, A., Omboga, Z., & Habwe, J. (2024). Perspective chapter: Court interpreters in the pursuit of sustainable development goals. In *Globalization and sustainability: Ecological, social and cultural perspectives*. IntechOpen.

Pierre, E. (2026). Backgrounding agency in EU institution instruments: A linguistic and argumentative perspective. *Argumentation and Advocacy*, 1–22.

Pomerantz, A. (1986). Extreme case formulations: A way of legitimizing claims. *Human Studies, 9*(2), 219–229.

Raineri, S. (2010). *Analyse contrastive français-anglais du passif dans une perspective constructionnelle : Sens et fonction de BE Ven, ÊTRE Vé, GET Ven et SE FAIRE Ver* [Doctoral dissertation, Université Sorbonne Nouvelle – Paris III].

Shibatani, M. (1985). Passives and related constructions: A prototype analysis. *Language, 61*(4), 821–848.

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition*. Blackwell.

Szabolcsi, A. (2010). *Quantification*. Cambridge University Press.

Trope, Y., & Liberman, N. (2000). Temporal construal and time-dependent changes in preference. *Journal of Personality and Social Psychology, 79*(6), 876–889.

Trope, Y., & Liberman, N. (2010). Construal-level theory of psychological distance. *Psychological Review, 117*(2), 440–463.

Turney, P. D., & Littman, M. L. (2003). Measuring praise and criticism: Inference of semantic orientation from association. *ACM Transactions on Information Systems, 21*(4), 315–346.

Wilson, T., Wiebe, J., & Hoffmann, P. (2009). Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis. *Computational Linguistics, 35*(3), 399–433.