Preregistration:

1. Have any data been collected for this study already?
2. What’s the main question being asked or hypothesis being tested in this study?

* What factors influence the “se faire” vs. “être” passive alternation in French?
* In other words: what factors influence speakers’ choice between the “se faire” and the “être” passive construction?
* Operationalized as: the probability that the “se faire” passive construction is used

1. Describe the key dependent variable(s) specifying how they will be measured.

* DV: type of passive construction (“se faire”/“être”)
* Values: 0 = “être” construction; 1 = “se faire” construction
* Annotation criteria: …

1. How many and which conditions will participants be assigned to?

* This is a corpus study, so there will be no participants.
* Factors for the logistic regression:
  + Subject
    - Animacy of the subject (inanimate/animate)
    - Person of the subject (...)
    - Number of the subject (…)
    - Gender of the subject? (...)
  + Verb
    - Aspect/ Form of auxiliary verb ? (…)
    - Tense ? (…)
    - Main verb semantic group (clusters)
    - Telicity of the verb (telic/ atelic)
    - Adversativity of the verb (adversative/ neutral) or sentiment analysis values
  + Medium? Register? (informal/formal)
  + Complément d’agent (absent/present)
  + Periphrase constructions (present/absent)
  + Mode (spoken/ written)
  + Negation (present/absent)
* Interactions tested:
  + Animacy of subject and Complément d‘agent ?
  + Adversativity and Subject Animacy (adversative verbs with animate subjects) -> showing affectedness
  + …

1. Specify exactly which analyses you will conduct to examine the main question/hypothesis.

* Logistic regression
* The model will calculate the probability that a given construction is realized with „se faire“
* The predictor variables will be normalized (log of the odds)
* Random effects (to avoid individual effects):
  + lemma of the verb
  + corpus
  + person (metadata)?
* p-value 0.05?

1. Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.

* Constructions will only be included if:
  + the construction can be realized with both “se faire” and “être” passive
  + the verb is a transitive verb
  + tense of the verb? (temps composé vs. présent)
  + negations / complex phrases?
  + Se faire construction has a passive (and not an exclusively causative) reading
* ...

1. How many observations will be collected or what will determine sample size?
2. Anything else you would like to pre-register?