

French Subject Doubling: A Third Path

Yiming Liang^{1,2,3} Caterina Donati^{1,2,3} Heather Burnett^{1,2,3}

¹Laboratoire de Linguistique Formelle ²Université Paris Cité

³Centre National de la Recherche Scientifique

1. Introduction. This paper revisits the status of subject clitics in Spoken French and proposes a new analysis of subject doubling. Based on new evidence from a corpus study, we argue for an analysis reconciling two competing views of the construction, and provide further evidence of how quantitative studies of language use can shed light on long-standing theoretical debates.

2. The debate. Subject doubling, where a nominal subject and a coreferential subject clitic co-occur (*Marie_i elle_i mange* ‘Mary_i she_i eats’) is a common feature of spoken French. Two analyses have been proposed to account for the structure. Researchers like Kayne (1991), Rizzi (1986), De Cat (2005) a.o. analyze it as a case of **dislocation**, where the NP subject is dislocated into a topic position in the left periphery, while the subject clitic is a syntactically argument-bearing pronoun merged in SpecTP and phonologically cliticized to the verb (cf. (1a)). Evidence for this analysis comes from observations such as: 1) in elicitation studies, the subject clitic does not systematically co-occur with a nominal subject, hence behaving like an argument; 2) other clitics, like the negation particle ‘ne’, or object clitics can intervene between the subject clitic and the verb, showing their relative autonomy; 3) it is difficult for the DP to receive a focus reading when the subject clitic is present; and 4) subject doubling obeys the topicality hierarchy: it appears to be incompatible with indefinites and quantified noun phrases, which is expected under a dislocation analysis, since the DP would be a topic (Rizzi 1986).

(1) a. Dislocation: $[_{TOP} \text{Marie} [_{TP} \text{elle} [_T \text{mange}]]]$

b. Morphological: $[_{TP} \text{Marie} [_T \text{elle-mange}]]$

Other researchers, in particular Roberge (1990), Auger (1995, 2003), and more recently Culbertson (2010) a.o., argue for a **morphological** analysis, according to which the DP subject occupies the canonical subject position and the subject clitic is an agreement marker base-generated in T (cf. (1b)). Evidence for this analysis includes: 1) in many corpus studies of spoken French, subject doubling is nearly categorical (over 70%), as is the absence of *ne* in negative sentences (under 5%); 2) regardless of whether it is followed by a subject clitic, no phonological or prosodic features single out the subject DP as being dislocated; and 3) the subject clitic is acceptable when the sentence is in broad-focus contexts, showing that DP subject is not necessarily interpreted as a topic (Culbertson 2010). In order to account for subject doubling’s restriction to definites, Culbertson (2010) proposes that French subject doubling is subject to Suñer (1988)’s “matching hypothesis”, whereby agreement morphology and its argument controller must match featurally. Since subject clitics have the feature [+definite, +accessible], their DP controller must also bear these features.

Although some arguments have come from linguists’ intuitions or experiments, corpus studies have played an enormous role in the development of the morphological approach (ex. Auger, Coveney 2003, Culbertson, a.o.). This being said, previous quantitative work has only investigated a subset of the relevant aspects of this complex and puzzling linguistic phenomenon. For example, variationist sociolinguistic studies (Auger, Coveney, Zahler 2014) investigate both social factors and grammatical factors; however, they do not take into account more cognitive factors like informativity and frequency. Culbertson has a more cognitive perspective, but the grammatical contrasts that she studies are quite broad, not distinguishing, for example, between different kinds of declarative clauses. In order to get a fuller picture, we therefore decided to track the contours of this phenomenon in one of the most recent corpora of Spoken French: the *Multicultural Parisian French* corpus (Gadet & Guerin 2016).

3. Subject Doubling in the MPF. We extracted from the corpus the occurrences of nominal subjects (e.g., *mon père* ‘my father’, *Marie*) or quantified subjects (*certaines* ‘certain people’, *personne* ‘nobody’), and annotated whether the subject is doubled by a subject clitic or not (N=3903). In our first analysis, we only considered preverbal DP subjects. Strong pronouns were also excluded. In order to obtain the most complete corpus study of subject doubling to date, we coded all the extracted tokens for the following factors: social factors (age, gender, education, profession), the polarity of the clause (i.e., affirmative, negative with ‘ne’, negative without ‘ne’), verb frequency, DP subject head frequency, the distance (in words) between subject and verb, subject type and clause type (root, subordinate, relative). Speaker and verb lemma were added to the model as random intercepts.

Mixed effects logistic regression analysis of our data reveals significant effects that provide support both to the morphological approach and the dislocation analysis. Supporting the morphological approach: we find a very high rate of subject doubling (74%) and an extremely low rate of doubling with *ne* (0%). In fact, in a secondary analysis, we included postverbal subject doubling cases (e.g., *Il est là mon père*), and the rate of doubling increased to 77%. Furthermore, within the 23% of occurrences without subject clitics, most of them involve ‘ne’ or are about education and religion, signs they come from the formal Standard French register. We also find that verb frequency is positively correlated with subject doubling ($z = 3.201, p < 0.01$), which, we argue, provides indirect evidence in favour of the morphological approach, since high frequency is well known to preserve complex morphosyntactic forms (Bybee, 2003). Supporting the dislocation analysis: we find that doubling is governed by subject type in terms of topicality hierarchy: definite NPs (rate of doubling: 80%) > indefinite NPs (47%) > universal QPs (8%). This being said, recall that Culbertson (2010), who finds the same effect, does not consider this constraint evidence of dislocation, but is rather the product of matching. However, we argue that agreement matching is not sufficient to explain another significant effect arguing for a topicalized DP: clause type. We find that root clauses, including root interrogatives, are associated with the highest rate of subject doubling; whereas subordinate clauses disfavor it. Among them, relative clauses disfavor doubling most, while other subordinate clauses are in-between (root (77%) > other subordinates (63%) > relatives (33%); differences between two adjacent categories being significant $p < 0.001$ for both subject type and clause type). The matching hypothesis has nothing to say about how clause structure would affect doubling; however, an analysis in which the subject DP is some kind of topic does, since it is well known that there are topics (eg. Cinque 1977’s “hanging topic”) that are only available in root clauses (cf. Albrecht et al. 2012, a.o.).

4. New proposal. To account for both the overwhelming evidence that Spoken French subject clitics are agreement markers, and the evidence that subject DPs are topics in doubling construction, we propose an analysis that is a hybrid of the dislocation and morphological analyses. We propose that subject doubling involves an agreement marker generated in T which is doubled by a DP located in low topic position which is only present in root clauses (2).

(2) Our analysis: $[_{TOP} \text{Marie} [_{TP} \text{pro}_T \text{elle-mange}]]$

Our analysis correctly predicts that the subject DP will obey the topic hierarchy and that doubling will be strongly disfavoured with relative clauses, since the subject DP acts as an intervener for the relative A-bar dependency. A corollary of our analysis is that the informal register of spoken French is a null-subject language, as proposed by Roberge (1990), Culbertson (2010) and others, and claimed for other subject doubling Romance languages like Picard (Auger) or Northern Italian dialects (Poletto 2000). Our analysis of the subject clitic as an agreement morpheme coupled with a topic-like preverbal subject indeed echoes Alexiadou & Anagnostopoulou (1998)’s hypothesis on null subject languages.

Selected References. 1. Aelbrecht, L., Haegeman, L., & Nye, R. (2012). *Main Clause Phenomena: New Horizons*. John Benjamins. 2. Alexiadou, A., & Anagnostopoulou, E. (1998). Parametrizing AGR: Word order, V-movement and EPP-checking. *Natural Language & Linguistic Theory*, 16(3), 491–539. 3. Auger, J. (1995). Les clitiques pronominaux en français parlé informel: Une approche morphologique. *Revue québécoise de linguistique*, 24(1), 21–60. 4. Culbertson, J. (2010). Convergent evidence for categorial change in French: From subject clitic to agreement marker. *Language*, 86(1), 85–132. 5. De Cat, C. (2005). French subject clitics are not agreement markers. *Lingua*, 115(9), 1195–1219. 6. Kayne, R. S. (1991). Romance clitics, verb movement, and PRO. *Linguistic Inquiry*, 22(4), 647–686.