Journal of Pragmatics xxx (xxxx) xxx

Contents lists available at ScienceDirect

Journal of Pragmatics



journal homepage: www.elsevier.com/locate/pragma

Topic chains in dialogues

Aria Adli

University of Cologne, Germany

A R T I C L E I N F O

Article history: Available online xxx

Keywords: Topic chain Domain Referential chain Information structure Null subjects Discourse coherence

ABSTRACT

This paper analyzes topic chains (TChs) in spontaneous speech dialogues. The chainbuilding property of topical elements is an essential means for managing prominence on the discourse level, creating thematic coherence across sentences. TChs relate the sentential and the discourse aspect of information structure by extending the sentenceinternal division between prominent and non-prominent information to the crosssentential level, where we can distinguish between categorical sentences that continue a TCh and those that begin a new TCh. The paper examines the role of the speaker and addressee feature (Harley and Ritter, 2002) on chain building. My assumption is that prominence management differs between 3rd and 1st/2nd person topics, i.e. according to [-local] or [+local] (Ritter and Wiltschko, 2009). I assume that common assumptions on TChs (Chafe, 1976; Reinhart, 1981; Brunetti, 2009) are plausible with 3rd person referents, but much less so with local persons: 1st and 2nd persons are never new referents in a dialogue situation. I therefore propose an analysis with two parallel TChs: one in the local domain and one in the nonlocal domain. The paper concentrates on topical overt and null subject pronouns in Spanish. I will demonstrate that a detailed analysis of topicality differentiating between two parallel chains allows us to understand hitherto unexplained variance with regard to the variation between sentences with realized subject pronouns (e.g. yo canto esta canción 'I sing this song') and their counterparts with zero subjects (e.g. Ø canto esta canción 'I sing this song').

© 2019 Published by Elsevier B.V.

1. Introduction

All comprehensible discourse (narratives, natural dialogues, etc.) is a structured entity which contains a coherent, and thereby interpretable structure of referential expressions. These traverse the text as a network of referential chains. Many entities (individuals, objects, abstract notions) referred to in a text are newly introduced at some point and then taken up again, forming a referential chain. This paper deals with a subset of referential expressions, namely topics (see section 1.1). The following passage (taken from the Spanish part of the sgs database) shows one topic chain (TCh) from (1a) to (1b) and another one from (1c) to (1d).¹ While referential chains describe a property of text structure, topics describe a property of common ground (CG) management and bring in the speaker/hearer perspective, which is particularly important to dialogue:

https://doi.org/10.1016/j.pragma.2019.07.022 0378-2166/© 2019 Published by Elsevier B.V.



E-mail address: aria.adli@uni-koeln.de.

¹ In the discourse extracts discussed in this paper, all referential expressions that are topic in at least one sentence of the extract have an index. A selection of referential expressions that are not topic, those which are human, are also indexed. Please note that referential (and thereby also topic) chains are independent of the speaker of the sentence.

ARTICLE IN PRESS

A. Adli / Journal of Pragmatics xxx (xxxx) xxx

The function of the topic for speaker and hearer is to "identify the entity or set of entities under which the information expressed in the comment constituent should be stored in the CG content" (Krifka, 2007: 41). Entities present in the communicative scene, in particular discourse participants, differ from other entities, because they do not need to be introduced due to their continuous situational activation. The question I pose in this paper is whether the dissimilar cognitive activation status of discourse participants on the one hand and all other speaker/hearer-external referents on the ot is reflected in disparate CG management. More specifically, I ask whether there are two discernible types of TChs that span the sentence topics of a discourse or whether there is a single TCh regardless of grammatical person.

(1a)	¿Y cór	no estaba	[_{Top} el arcón] _p	?	
	and how	wwas.3se	[the freezer]]p	
	'And how w	vas the freezer?'			
(1b)	ί[τοp pro]p	estaba ce	rrado? ¿[Top pro]p	estaba	abierto?
	prop	was.38G clo	osed.M pro _p	was.3SG	open.M
	'Was it clos	sed? Was it open	?'		
(1c)	¿Cómo	estaba	[_{Top} ella] _q ?		
	how	was.3SG	she _q		
	'How was s	she?' (17/69)			
	/				
(1d)	¿[_{Тор} <i>р</i> ро] _q	estaba	encogida?		
	pro_q	was.3SG	shrunken.F		
	'Was she sh	nrunken?'			

By employing this distinction, I aim at a better understanding of the topic concept in general. The general goal is to explore the adequacy of a double TCh model that I propose in this paper. We will evaluate the adequacy of a model in which we assume two parallel TChs, one for 3rd person referents called TCh in the nonlocal domain, and one for the actual discourse participants called TCh in the local domain (locality refers to the participants). This means that the speaker stores two types of topic in the course of the discourse, but as sentences typically only allow for one topic (from the local or nonlocal domain), we only find one of them in a particular sentence. In sections 1.1 and 1.2 I lay out my definitions of TChs and the topic itself, as groundwork for the analysis of double chains later on.

1.1. Chains, domains and features



A. Adli / Journal of Pragmatics xxx (xxxx) xxx

1.1.1. Topic chains and referential chains

If we modify the discourse extract by inserting sentence (1c') with a local subject topic between (1c) and (1d), we obtain an interesting scenario (TChs are visualized by continuous lines, (non-topical) referential chains are visualized by broken lines). Under a single TCh analysis, there would be no TCh connecting 'ella' in (1c) and the null subject in (1d), whereas there is one (as shown above) under the scope of a double TCh analysis. This example also serves to illustrate the relation between referential chains and TChs: First, the presence of a referential chain does not imply the presence of a TCh. Second, if there is a TCh, it can only include elements of a referential chain, but not necessarily all of them. Thus, the expressions of a TCh are a subset of the expressions of a referential chain starts with the first and ends with the last expression in the text referring to the same referent. Intervening expressions referring to other referents do not 'break the chain'. A TCh links all sentence topics that refer to the same referent. Unlike referential chains, TChs end when a topic expression intervenes that refers to another referent. If later-on the sentence topic refers to the same referent, the topic expression would be considered to be part of a new TCh. Since chains are links, we need at least two expressions to form a referential or topic chain. In the remainder of the paper, the expression "chain" refers to topic chains (TChs), unless I specifically employ the term "referential chain".

1.1.2. Topic shift and domain switch

Whenever the topic shifts to another referent, a new TCh starts. Therefore, the first topic of a chain is called shift topic, while the following ones are called familiar topics ('familiar' in the sense that the topic referent is the same as the previous topic referent in the discourse). Under the assumption of a single TCh that does not distinguish between discourse participants and other referents, the 2sg null subject pronoun at the beginning of (1c') interrupts the TCh started in the previous sentence (1c). We have to analyze the null subject in the following sentence (1d) again as a shift topic, i.e. we have one shift topic in (1c), one shift topic in (1c), and again a shift topic in (1d). However, by utilizing a double chain model, the insertion of (1c') concerns the TCh in the local domain and does not modify the TCh in the nonlocal domain that links (1c) to (1d). Rather, we have a (topic) domain switch in (1c') from the TCh in the nonlocal domain to the one in the local domain. And in (1d) the domain switches again, this time from the local domain back to the nonlocal domain. In short, a topic shift signals a change in the topic entity within the same chain, while a domain switch signals – under a double chain model - a change of the presently forwarded TCh (either in the local or nonlocal domain). Topic shift and domain switch can occur independently of each other. The following three discourse extracts, in which the sentence topic is either a certain dog, a certain cat, or one of the speakers, show all four combinations of [±shift] and [±switch]. The topic of sentence (2b) is a case of [-shift] and [-switch], since it is coreferential with the topic of the immediately preceding sentence (2a). The topic in sentence (2c) is [+shift] [-switch], because the topic shifts to the referent "cat", which still belongs to the same chain. In the next discourse extract, (3d) contains a case of a [-shift][+switch] topic: It is [-shift], because the previous topic entity of the same chain has already been the referent "dog" in (3a). At the same time, the active TCh switches from the one in the local domain in (3c), where "speaker A" has been topic, to the one in the nonlocal domain in (3d). Finally, (4e) shows a [+shift][+switch] sentence topic. The previous topic entity of the same chain was a different referent, namely "dog" in (4b), and it shifts to "cat" in (4e). At the same time, the active TCh switches from the one in the local domain in (4d) to the one in the nonlocal domain in (4e).



1.1.3. Topics in the local and nonlocal domain

Using Reinhart's (1981) metaphor, we can imagine the double chain model as two piles of file cards that are next to each other. However, the file card for the local domain contains only a few cards, two in a dialogue. If one considers them as two separate piles of file cards, sentences with a local domain topic cannot interrupt the continuity of a chain in the nonlocal domain that might be present in the immediately preceding and following context. While inserting occasional questions or comments with a 1sg or 2sg pronoun in topic position in dialogues speakers often talk about some external referent across a series of sentences. The same can be said about monologues with regard to 1sg. In such cases, one can model a higher degree of discourse coherence by sentence topic continuity, without resorting to the notion of discourse topic (Roberts, 2011). A double chain more accurately takes into account the observation that these occasional insertions of questions/comments with

ARTICLE IN PRESS

A. Adli / Journal of Pragmatics xxx (xxxx) xxx

1sg or 2sg topic do not lower discourse coherence, especially in comparison with insertions of utterances with 3rd person shift topics.

From a semantic point of view referents in the local domain differ from those in the nonlocal domain due to their deictic nature. It is an issue of debate whether or not discourse participants are, on a par with referents in the nonlocal domain, anaphoric expressions. I will not dive into this debate here. For the purposes of this paper it is sufficient to state that the function of deictic personal pronouns is not mere identification of the discourse participants, but the definition of their semantic roles in the event expressed in the sentence (Diessel, 2012: 2415 points out that "the use of person deictics is similar to the use of anaphors"). Therefore, I assume that deictic personal pronouns can function as topics.

1.1.4. Feature representation of topic chains

The different properties of a single- and double chain can be illustrated by features and their combinations: Under a single-chain model there is only one chain-related feature with two values, namely [\pm shift]. When employing a double-chain model each sentence topic is specified for three chain-related features: [\pm shift] indicating whether it is the start or the continuation of a TCh, [\pm local] indicating whether it belongs to a chain of the local or to a chain of the nonlocal domain, and [\pm switch] indicating whether the present topic belongs to a chain of the other domain than the previous topic referent (which would be [+switch]) or of the same domain (which would be [-switch]). In consequence, the double-chain model distinguishes eight different feature combinations, four regarding local domain chains and four regarding nonlocal domain chains. In a dialogue, [+shift] within the local domain indicates topic shifts from one speaker to the other, while [+shift] within the nonlocal domain is not limited to a pre-established closed set of entities.

	[+]0	ocal]		[-local]				
[+sw	vitch]	[-sv	vitch]	[+sw	vitch]	[-switch]		
[+shift]	[-shift]	[+shift]	[-shift]	[+shift] [-shift]		[+shift]	[-shift]	

The double-chain model will result in fewer instances of shift topics, because familiar topics in domain-switching contexts, namely [+switch][-shift][+local] and [+switch][-local] (e.g. in (1d)), would be analyzed as shift topics in the single-chain model. Importantly, an annotation of all topics based on the three features [±shift], [±local], and [±switch] allows us not only to represent the discourse structure with a double chain model, but also the depict the discourse structure with a single chain model, because we can directly derive the single chain [±shift] feature from the three double chain features.

1.2. Identifying topics in natural dialogues

The very notion of topic is a notoriously difficult one within academic debate. Although (or *because*) its conceptual definition is fairly intuitive and straightforward ("what a sentence is about" or "the entity under which the information expressed in the comment is stored in the CG") its operational definition that allows us to decide exactly which of the expressions in a sentence is topical, is difficult up to the point of frustration. The conceptual definition leaves room for a variety of operationalizations. In the following, I propose an operational definition building on the *basic assumptions on the grammatical realization of topics* in (5). (5a) to (5c) specify or restrict the 'search domain' in a sentence within which the topic can occur. (5d) to (5g) more narrowly define grammatical and semantic conditions of topichood:

- (5a) Elliptical sentences can contain a topic, which is typically located in the elided part.
- (5b) (Non-contrastive) topics do not occur in the focus domain but only in the backgrounded part.
- (5c) Topics are root phenomena in the sense that they only occur in the main speech act. Consequently embedded clauses do not contain topics unless the main speech act is embedded.
- (5d) Only verbal arguments and not adjuncts/modifiers can be topic. Framesetters (that are a feature of certain categorical sentences, see Jacobs, 2001: 655–658) and stage topics (that are typically modifiers) are not to be misrepresented by employing the topic notion.
- (5e) In most sentence types, the subject is the prototypical topic candidate. However, in certain grammatical structures, an object or PP complement can be the prototypical candidate.
- (5f) The semantic properties of an expression can disfavor its use as topic. In such cases, another argument might become topic.
- (5g) Sentences can be topicless. Either because of a particular grammatical/prosodic structure that makes them thetic. Or because the topic candidate is focused or does not meet the semantic criteria.

A. Adli / Journal of Pragmatics xxx (xxxx) xxx



Fig. 1. Algorithm for topic annotation.

The operational definition based on these basic assumptions has been translated into the *algorithm* in Fig. 1 (represented by a flowchart according to the ISO 5807:1985 standard). It goes through a text and takes the different criteria into account in a meaningful order. Step 1 is the iterative loop that proceeds sentence by sentence through the text. In step 2, we ask whether the current sentence is thetic, in which case one does not need to proceed any further. An example of a thetic sentence would be (6).

(6) (What happened to you?) My DOG got sick.

ARTICLE IN PRESS

A. Adli / Journal of Pragmatics xxx (xxxx) xxx

A thorough overview of thetic structures is given by Sasse (2006) (he highlights up to four distinctive constructions in European languages: subject accentuation, verb-subject order, split subject+relative clause structures, and subject incorporation).

Step 3 and 4 identify ellipses – a frequent phenomenon occurring in spontaneous dialogue. As a rule the elided part is reconstructed by means of the immediately preceding context. As demonstrated in (7b), the elided part, represented through the use of round parentheses, typically contains the entity that continues the TCh. More examples can be found in (25c)–(25f) in section 3.4.

- (7a) A: How long did [Top he] speak with Mary?
- (7b) B: ([_{Top} He] spoke with her for) 1 hour.

Step 5 checks whether the sentence contains a contrastive topic in the sense of Krifka's (2007: 47–48) notion of delimitation, where the speaker signals a partial or sequential answer. If this is the case, the search is concluded with the identification of the contrastive topic.

Step 6 analyzes another information-structural dimension, namely the focus-background partitioning of the sentence. This is indispensable, because topics are only possible in the background part (with the exception of contrastive topics which have a $[_{Top} [_{Foc}]]$ structure). All entities in the focus domain are thereby excluded as viable topic candidates. If the background part contains no argument, the sentence is topicless although it does not automatically need to correspond to a typical thetic construction. One case in point is (8), in which the subject 'Peter' is the only argument and narrowly focused.

(8) (Who slept on the couch last night?) [Foc Peter] did.

Step 7 takes into account complex sentences which require closer examination with regard to topicality. Certain transitive matrix verbs – that are however frequent among complex sentences in spontaneous speech – are functionally similar to parentheticals. To simplify matters, this is the case in point with verbs of saying, thinking/believing, knowing (see Hooper and Thompson, 1973 for a more complete list). The main assertion is the complement clause (see also Bianchi and Frascarelli, 2010; Krifka, 2014). In these cases (and only in these cases) the topic entity is to be searched for in the complement clause instead of the matrix clause, as shown in (9). Other examples are (22c) or (25c) in sections 3.3 and 3.4.

(9) So you think that [Top they] were angry with him.

Step 8 shortens the algorithm in cases of dislocated entities, which are unequivocally topics. In some languages dislocations can (rarely) create multiple topic constructions, as the French example (10) demonstrates.

(10)	Fr. [_{Top} Toi],	[_{Top} le voleur],	tu	ľ	as	vu?
	You	the thief	you	him	have.2sg	seen

Step 9 deals with the relationship between topic and argument structure. Arguments with a specific syntactic function are preferred as topics, which we refer to as *prototypical topic candidates according to syntax*. In most constructions, the subject is preferred as topic, e.g. the subject *he* in (7b) is a prototypical candidate while the complement *her* is non-prototypical. However, a non-subject argument can also be the prototypical candidate (see Brunetti, 2009), if it is what is a so-called 'logical subject'. E.g. the dative experiencer in sentences with a psychological verb as in (11) is favoured over the grammatical subject as topic. A dislocated argument overrides step 9 because it is always the prototypical candidate, irrespective of its syntactic function.

(11) Sp.[_{Top} Le] gusta Michael Jackson him/her.dat like.3sg Michael Jackson.nom

A. Adli / Journal of Pragmatics xxx (xxxx) xxx

Step 10 assesses the semantic topichood criteria of the prototypical candidate. Features assumed to be crucial for topicality are animacy, agentivity, definiteness, specificity, genericity. This list could probably be extended. Unless the prototypical candidate does not meet pre-defined minimal criteria, it is confirmed as topic. Otherwise, the other arguments are considered, but in this case, they need to fulfill (near-) optimal criteria, e.g. being human, definite, specific, and given, in order to qualify as topic. For example, if the subject in (9) was indefinite and non-specific as in (12), 'him' would become topic. Another case in point is demonstrated in (22c) '...some neighbor has ever gone to any of the parties?', which is topicless because no argument meets the semantic criteria.

(12) So you think that someone was angry with [Top him].

Subsequently non-subject topics are rare, but they can occur if a non-subject is dislocated, or if a non-subject occurs in a construction in which it is the 'logical subject', or if the subject fails the semantic criteria while a non-subject argument fully meets them.

The idea of a ranking of topic candidates is reminiscent of the ranking of forward-looking centers C_f in Centering Theory (Grosz et al., 1995), where the referring expressions in a sentence are ranked. However, there are some important differences: First, Centering Theory concentrates on adjacent sentences, while our chain-based model allows for more global relations (we come back to the question of distance in the discussion in section 4). Second, the topic algorithm above leads to a clear prediction on the actual sentence topic, while the forward-looking centers, being partially ordered, only allow for a probabilistic prediction (on the most likely backward-looking center in the following sentence). Third, the topic algorithm takes into account a variety of syntactic, semantic, and pragmatic factors. Fourth, there is no direct correlate for the backward-looking center in this topic analysis, because the topic of a given sentence does not per se maintain a particular information-structural status in the following sentence where it can again occur as topic or occur in the comment part (if it occurs) (see also Hoffman, 1998 on the different roles of information structure and centering in discourse processing). This being said, the choice of referring expressions in discourse is relevant in both Centering Theory and to our topic analysis in which we correlate null and overt subject pronouns with topic continuity and shift.

2. Corpuslinguistic, variationist approach

Many null subject languages show a correlation between grammatical form of the subject and topic continuity (more precisely, consistent null-subject languages in the sense of Holmberg, 2005; Holmberg et al., 2009): When a subject pronoun is topic, its probability to be realized as a null form is higher if it is a familiar topic, i.e. continuing a TCh. Likewise its probability to be realized overtly is higher if it is a shift topic. If we take another look at the initial discourse segment, a new TCh is established by the subject of the first sentence in (1a) (arcón 'freezer'), and continued by the two null subjects in the following sentences in (1b), which are familiar (henceforth [-shift]) topics. Then, the overt pronoun ella 'she' establishes a new TCh in (1c), i.e. it is a shift (henceforth [+shift]) topic. The null subject of the ensuing sentence (1d) continues the chain, which is again a [-shift] topic. This is a matter of tendency or probability (Adli, 2011) and not a categorical or hard grammatical principle (as has been claimed for example by Frascarelli, 2007). The correlation between grammatical form of the subject and topic continuity can be counted in a corpus and expressed by the proportion of overt pronouns among all referential pronouns, which is a value between 0 and 1 called pronoun rate (see section 2.2). This correlation has no influence on the question of how TChs are construed in discourse. However, it offers a useful criterion for evaluating the double chain model. For example, under the assumption of a single chain model the insertion of (1c') should raise the probability that the subject pronoun in (1d) is realized by the overt form *ella* instead of null. Under the assumption of a double chain model on the other hand, the insertion of (1c') should have no effect on the grammatical form of the subject pronoun in (1d).

The data has been prepared in a two-step procedure: First, I identified *all* topics with the three chain-related features [±shift], [±local], and [±switch], i.e. I established the chain structure in the discourse. Then, I selected a *subset* of sentences, in which the topics are *subject* and *pronominal*, in order to be able to work with a quantitative criterion. Please recall that topics can occur as the subject and sometimes also as a non-subject argument. And they can be pronominal or full lexical expressions. A TCh can contain all the possibilities. In the following we will analyse the effect of the three features [±shift], [±local], and [±switch] one by one and how they interact. This is achieved by visualizing the pronoun rate (for the subset of pronominal subject topics) for different feature combinations in bar charts. Whenever I see noteworthy results in the charts that could hint towards an effect, I will scrutinize them by selecting illustrative passages from the corpus, in order to gain a better understanding of the discourse patterns behind. The entire approach in this paper is, terms of empirical methodology, aimed at generating and not testing hypotheses. I provide a systematic report of the frequencies, but do not apply inferential statistical tests. Given that the idea of a double chain is new, the aim of this study is a *proof of concept*. Therefore,

7

ARTICLE IN PRESS

A. Adli / Journal of Pragmatics xxx (xxxx) xxx

the focus in section 3 lies in exploring and discussing discourse extracts that can explain some of the results shown in the bar charts. The decision in favor or against the double chain model should build on the usefulness of this modelization of topic continuity, i.e. whether it helps us to uncover and understand systematic discourse patterns in a meaningful way or not.

2.1. The spontaneous speech data source

I use Spanish data from the sgs corpus (www.sgscorpus.com), consisting of spoken dialogue data that has been collected by the author during fieldwork in Barcelona in the year 2008 with native speakers of the Spanish variety of Catalonia. Catalonia is a bilingual region, I selected balanced bilinguals that speak both Spanish and Catalan since early childhood in their daily life. In order to be able to generalize the results to a certain extent, I work with a sample of speakers: The data consists of transcribed and annotated recordings of 54 persons who were selected according to a given sample plan. We tried to obtain a sample that is representative of a part of the population with the following characteristics: Adults of young and medium age (17–48 years, mean age 28), who had at least obtained compulsory education, most of them are two or more years above the compulsory education (Spanish Ciclos Formativos or Bachillerato). Furthermore, I aimed for a balance between male and female speakers (60% women, 40% men). Interviewees solved a fictive murder case, speaking freely with the (native and trained) interviewer. Mostly, interviewees chose a non-formal, colloquial style, encouraged by an initial icebreaker phase and the interviewer's style. Compared with the Labovian-style sociolinguistic interview, this data leads to a dialogue which is dynamic in the sense that the discourse participants often change turns, large monologues are absent. The Spanish data of sgs consists of 20,440 utterances, 10,778 (52.7%) by the 54 interviewees, 9,662 (47.3%) by the other participants, mainly by the two native interviewers. The interviewees' utterances consist of 6,195 (60.2%) non-elliptical and 1,692 elliptical (16.4%) constructions, the rest are phenomena that are excluded from any grammatical analyses such as false start, single interaction markers ('m-m', etc.). The complete, i.e. non-elliptical, sentences of the interviewees consist of 1,117 declaratives (18.0%), 4,993 interrogatives (80.6%), 24 exclamatives (0.4%), and 61 imperatives (1.0%), see Adli (2011: 32–33) for full details.

In a first step, the recordings were transcribed, segmented, and aligned with the audio. The discourse segmentation corresponds to the requirements of the topic analysis outlined in section 1.2: It is segmented into sentences, which correspond to one root clause with its subordinate clauses if there are any. This complex or simple sentence is also the unit that can contain a topic. Elliptical sentences and one-word utterances are also separate units. In a second step, the grammatical function of each major constituent (subject, direct object, indirect object, prepositional complement, predicate, different types of adjuncts, subjunction) and the category or grammatical form of the respective constituents has been annotated. Furthermore, in complex sentences the hierarchical structure of the different clauses, i.e. the clausal dependency relations, clausal coordination and the function of the subordinate clauses for their respective matrix clauses has been coded. I analyzed *all* utterances of the interviews (and not just a selected portion of 'unproblematic' or prototypical cases). TChs span over the entire discourse. I therefore believe that the results can be generalized to language use in spontaneous interpersonal communication.

2.2. Calculating pronoun rates for pronominal subject topics

The pronoun rates given in the bar charts below are calculated as in (13). They are the mean of each person's pronoun rate (the speaker-sample-value in Adli, 2011: 229). In essence, I divide the number of overt pronouns by the number of all referential (overt and null) pronouns. n is the number of interviewees, i.e. 54, N_{OVERTi} is the number of clauses with overt subject pronoun and N_{NULLi} is the number of clauses with null subject pronoun uttered by the *i*-th interviewee.

(13) mean pronoun rate
$$h = \sum_{i=1}^{n} \frac{N_{overt_i}}{N_{overt_i} + N_{null_i}}/n$$

Since my quantification focuses on the grammatical form of subject topics, I can only calculate the pronoun rate for those sentences that (i) have an overt or null subject pronoun which (ii) is topic, and in which (iii) variation is theoretically possible between the overt and the null form – the latter circumscribes the set to the variable context or envelope of variation (Labov, 1972: 72). I defined the envelope, essentially following the careful definition of criteria proposed by Otheguy et al. (2007: 775–778): I excluded infinite clauses, sentences with an impersonal verb such as *hay* ('there is'), subject-headed relative clauses (including headless relative clauses, sentential relative clauses, and genitive relative clauses), inanimate subjects, elliptical sentences, and of course sentences whose subject is not a referential overt or null pronoun (which also excludes sentential subjects).

A. Adli / Journal of Pragmatics xxx (xxxx) xxx

Most importantly, this does not mean that the sentences that are not included do not have an effect on the results. To the contrary, they are constitutive to the structure of the TChs across the dialogue. Once again: TChs are annotated in the entire discourse, regardless of the grammatical function of the topic, the referential form of the subject expression, or the possibility of variation between overt and null subject. Only the quantification of the relation between subject pronouns and topics is restricted to the envelope of variation. The envelope of variation as defined above contains 679 overt and 3,058 null subject topics, produced by the 54 interviewees (Adli, 2011: 237). The analyses build on a pool of 679 + 3,058 = 3,737 sentences that are then split according to the features [±shift], [±switch], and [±local].

3. Corpus analysis of chain features

In section 3.1, I briefly assess to which extent the number of [+shift] topics decreases in the double chain model compared to the single chain model, as outlined in section 1.1. In 3.2, I disentangle different aspects of the double chain model, always with regard to topic continuity: I start by contrasting topic chains in the local and the nonlocal domain, then scrutinize first of all the role of domain-switching, and then the combination of topic chains in local vs. nonlocal domain with domain-switching. In section 3.3, I distinguish within the local domain between 1sg and 2sg topics, and within the nonlocal domain between 3sg and 3pl topic pronouns. Particular attention is paid to domain-switching familiar topics, because they would be analyzed differently, as [+shift], in a single chain model. In 3.4, I assess the role of the immediately activated, preceding context in topic shifts. In sum, I analyze all the chain-related features, steadily increasing the level of granularity.

3.1. Comparing single and double chain model

		single chain model	double chain model
(14a)	¿Cómo estaba [_{Top} ella] _q ?	[+shift]	[+switch][+shift][-local]
	how was.3SG she _q		
	'How was she?'		
(14b)	¿[_{Top} pro] _b la _q has visto?	[+shift]	[+switch][±shift][+local]
	prob her _q have.2SG seen		
	'Have you seen her?'		
(14c)	¿[Top pro]q estaba encogida?	[+shift]	[+switch][-shift][-local]
	pro _q was.3SG shrunken.F		
	'Was she shrunken?'		

An annotation according to the double chain model results in less than half the number of shift topics (195 overt and 582 null subjects) as compared to the single chain model (407 overt and 1226 null subjects), while it results in an increase of nearly a third in familiar topics (484 overt and 2476 null subjects) than the single chain model (272 overt and 1832 null subjects). This result is not surprising, as the discourse fragment (1c), (1c'), (1d) above has shown – repeated as (14a) to (14c). We see that the single chain model leads to [+shift] for all three sentences, while the double chain model leads to [-shift] for (14c) (since (14b) is an added fictive sentence, I leave it undefined with regard to [\pm shift]). One should keep in mind that the entities identified as sentence topic are the same in both models – the difference is in their chain structure (or in their [\pm shift] feature).

3.2. Assessing features in the double chain model: [±shift], [±switch], and [±local]

The following bar charts show the results for different chain features. The absolute frequencies of null + overt pronouns, i.e. the number of sentences in the corpus with a pronominal subject topic and the relevant feature combination, are printed under each bar. The y-axis indicates the pronoun rate calculated as explicated in (13), which is also printed above each bar. For example, the leftmost bar in Fig. 2 tells us that 28% of subject pronouns that are shift topics in the local domain are overtly expressed (which means in turn that 72% of them are null forms). That bar also tells us that this percentage is calculated based on a total of 385 sentences. Please recall that the following types of more fine-grained analyses would not be possible in the single-chain model, where the concept of (topic) domain switching does not exist. I expect (i) that shift topics come with a higher pronoun rate than familiar topics. Furthermore, I expect (ii) that domain-switching by itself does not lead to a higher pronoun rate (or only indirectly to a slightly higher rate because of the increase in referential distance that comes with

A. Adli / Journal of Pragmatics xxx (xxxx) xxx

domain-switching). I will highlight especially those results that further elaborate or deviate from these basic predictions. They are summarized in (Observation 1) to (Observation 5) below. These results will bring to light some new aspects of reference continuity in discourse (see Fig. 3).



Fig. 2. Topic continuity in the local and nonlocal.

Fig. 2 shows topic continuity ([±shift]) in the local and nonlocal domain, with the pronoun rates for each of the four feature combinations. The main finding here is that the first two bars are nearly identical, i.e. the average rate of pronominal topic expressions that refer to the speakers themselves (i.e. that belong to the local domain) is essentially independent of a topic shift from one speaker to the other.

(Observation 1): The difference in pronoun rate between shift and familiar topics is hardly noticeable with topics in the local domain.

This difference in pronoun rate between the discourse participants and the external or nonlocal referents, where pronoun rate of shift topics is twice as high than familiar topics, is a first indicator that a distinction between these two sets of referents makes sense. It is important to recall that in a dialogue between two speakers, shift topics in the local domain only occur when the topic referent shifts from one speaker to the other, independently of who utters the sentences. In other words, change in reference (who the sentence is about) matters, not turn change (who speaks).

Fig. 2 exemplifies topic continuity in domain-switching and same-domain contexts and gives us a second indication that the double chain approach is a promising option for modeling TChs in discourse: Shift and familiar topics demonstrate a clear difference in same-domain contexts, but much less so in domain-switching contexts. Two findings merit further attention: The first noteworthy finding, (Observation 2), summarizes the comparison between bar 1&2 of Fig. 2. In domain-switching contexts shift topics do not have a higher pronoun rate than familiar topics – we even observe a minor difference in the opposite direction. The second remarkable finding, (Observation 3), is based on the comparison of bar 1&3, unveiling a peculiar difference between domain-switching and same-domain contexts for shift topics.



Fig. 3. Topic continuity in domain-switching and same-domain contexts.

A. Adli / Journal of Pragmatics xxx (xxxx) xxx

(Observation 2):In domain-switching contexts shift topics do *not* have a higher pronoun rate than familiar topics.(Observation 3):Shift topics have a *lower* pronoun rate in domain-switching contexts than in same-domain contexts.

In the remainder of section 3, I will consecutively increase the granularity of the pronoun rates given so far and present some discourse examples that illustrate the results. The finer-grained picture will result in a modification/ revision of (Observation 1) to (Observation 3). I discuss these results in light of concrete discourse examples and propose several new hypotheses in (Observation 4) to (Observation 5). Please bear in mind that the differences in pronoun rate we are dealing with are *tendencies* in terms of more or less overt pronouns. The different discourse examples I am presenting are not representative for all examples in the corpus. Rather, they illustrate, after a thorough sighting of the material, certain salient and rather frequent discourse structures. Also recall that the results in the bar charts are based on the subset of pronominal subject topics. The discourse extracts are chosen so that we see a target sentence with a pronominal subject topic and the respective combination of [\pm switch], [\pm shift], [\pm local], person/ number, etc. *within its discourse environment*, where we may also find non-pronominal subject topics and non-subject topics.

Let us begin with (Observation 3), which summarizes a peculiar property of shift topics. Fig. 4 distinguishes between topics belonging to the local domain (bar 1–4) and those belonging to the nonlocal one (bar 5–8). The finer-grained picture confirms this observation, both for the local domain (see bar 1&3) and the nonlocal one (see bar 5&7). In order to gain a better understanding of types of discourse patterns behind (Observation 3), we scrutinize two examples of shift topics, one in a same domain and one in a domain-switching context:



Fig. 4. Topic continuity in the local & nonlocal domain, in domain-switching & same-domain contexts.

3.2.1. Shift topics in same-domain contexts

Sentences (15a)–(15c) are a discourse fragment, in which (15c) represents a case of shift topic in a same-domain context ([+shift], [-switch], [-local]) and in which the pronominal subject topic is overtly realized. One factor that seems to lead to an overt realization is *ease of disambiguation* between two active topic referents in the nonlocal domain (*los padres* 'the parents' and the female victim, the latter refered to by *ella* 'she'), in particular when the co-occurrence of the referents is accompanied by a topic shift: In (15a), 'the parents' are topic and the female victim occurs in the comment part. (15b) also contains these referents, yet both are in the comment. (15b) is topicless, because the entire sentence is focused, being the answer to the *wh*-object in (15a). In (15c), the female victim is topic and the parents are in the comment part, i.e. an inverse distribution compared to the preceding categorical sentence (15a).

11

12	A. Adli / Journal of Pragmatics xxx (xxxx) xxx	
(15a)	 A: ¿y qué lep decían exactamente [Top los padres]q? and what herp said.3PL exactly [the parents]q 'And what exactly did the parents tell her?" 	
(15b)	B: que pro_{q} no querian que pro_{p} volviera a verlo _r . that pro_{q} not wanted.3PL that pro_{p} returned.3SG.SBJV to see:him _r	
(15c)	'That they did not want her to see him again.' A: $y [_{Top} ella]_p^{-} les_q$ respondia que no? and them answered 280 that no	
	And she answered them "no"?" $(33/404)$	

3.2.2. Shift topics in domain-switching contexts

The discourse fragment (16a)-(16f) also shows a topic shift, but one in a domain-switching context. The topic of (16f) is a 3sg null subject topic – shifting from another nonlocal topic referent (having a snake) in (16d) to the female victim in (16f). This shift occurs in a domain-switching context, because the topic in the immediately preceding sentence (16e) belongs to a chain in the local domain. What is important is that the direct object referent (the snake, refered to by the clitic *la* 'her') in the comment part of (16f) is the same as in the preceding sentence (16e).

Given that the snake is refered to by the object clitic in (16f), it is essentially excluded as a possible referent for the null subject. Similarly, the fact of having a snake, which is the previous topic in the nonlocal domain in (16d), is not among possible referents, either. Thus, disambiguation of the null subject topic in (16f) seems unproblematic.

In sum, these two discourse fragments suggest that presence or absence of a prominent competitor is an important factor. In the same-domain context, the topic referent in (15c) also occurs in the preceding sentence, while in the domain-switching context in (16f) the previous occurrence of the topic referent of the same domain is further distant and therefore not such a prominent competitor any more.

(16a)	B:	sí, $[T_{\text{top}} pro]_p$ tenía $[una \text{ serpiente}]_q$.
		yes $/pro_{p}$ had.3SG [a snake] _q
		'Yes, she had a snake.'
(16b)	A:	¿sý?
		Yes?
(16c)	B:	sí.
		'Yes.'
(16d)	A:	$\operatorname{quizá} [T_{op} \operatorname{eso}]_r \operatorname{lap}_p$ ha matado.
		maybe this _r /her _p has killed
		'Maybe this has killed her.'
(16e)	B:	no hombre!, $[T_{\text{top}} pro]_{s}$ nos la_{q} hemos encontrado en la vitrina.
		no man pro_s REFL her _q have.1PL found in the glass.cage
		'No man, we found it in the glass cage.'
(16f)	A:	$_{i}$ $[_{Top}$ $pro]_{p}$ no $[a_{q}]$ dejaba nunca suelta por el piso?
		ah pro _p not her _q left.3SG never free.F on the floor
		'Ah, she never left it free on the floor?' (51/322)

A. Adli / Journal of Pragmatics xxx (xxxx) xxx

The greater zoom level in Fig. 4 allows us to scrutinize another finding: First, we can see that (Observation 1) is an apparent effect. A comparison between bar 1&2 and bar 3&4 shows that there *is* a difference between shift and familiar topics. However, in domain-switching contexts (bars 1&2) the direction is *opposite* to what we expect. This leads us to a revision of both (Observation 1) and (Observation 2), summarized in (Observation 4). Discourse examples will be given in section 3.3.

(Observation 4): In contexts of domain-switching to the local domain, shift topics have a *lower* pronoun rate than familiar topics.

So far, the empirical picture shows that domain-switching has a salient effect, but a different effect from topic shift (recall that under a single-chain model, domain-switching would always be viewed as topic shift). It lowers the effect of topic shifts on the average pronoun rate, see (Observation 3). And, in domain-switching to a chain in the local domain, it results in the even more surprising fact that shift topics have a lower pronoun rate than their familiar counterparts, see (Observation 4).



Fig. 5. Topic continuity according to PN in domain-switching and same-domain contexts.

3.3. Splitting up by person & number

In the following, I will divvy up topic chain in the local and nonlocal domain by person and number: With regard to the local domain I analyze 1sg and 2sg references separately (1pl and 2pl references are not included in this picture).² The nonlocal domain is separated into 3sg and 3pl references. The results are given in Fig. 5. It is important to highlight that the difference between the local and the nonlocal domain cannot simply be reduced to differences in person or number, which can also be represented in a single-chain model. The distinction between domain-specific chains is an architectural choice with far reaching implications for the depiction of how chains are construed. Please recall that e.g. a 1sg shift topic in a single-chain model.

We first turn to a comparison between familiar topics in domain-switching contexts and shift topics. As has been highlighted in section 1.1, they would all be treated alike as shift topics in a single-chain model. The results illustrate that familiar topics in domain-switching contexts do behave differently from shift topics: They have a lower pronoun rate for 1sg, but a higher pronoun rate for 2sg. With regard to 3sg, they have a lower pronoun rate than the average for shift topics (they are equal to shift topics in domain-switching contexts but have a lower rate than shift topics in same-domain contexts). With regard to 3pl, they also have a slightly lower rate. These findings support the view that a double-chain model allows the identification of some systematic differences that would otherwise have stayed unnoticed.

² 1pl references (7 shift and 70 familiar topics) and 2pl references (6 shift and 70 familiar topics) were too scarce in the corpus, especially for overt pronouns, to allow meaningful calculations of relative frequency, with a fine-grained resolution as in Fig. 5. However, future research on [±local], would require a distinction between inclusive and exclusive future references and discuss the hybrid status of exclusive references with regard to [±local].

ARTICLE IN PRESS

A. Adli / Journal of Pragmatics xxx (xxxx) xxx

Given the noteworthy picture unveiled in section 3.2, summarized in (Observation 3) and (Observation 4), the finergrained picture illustrated by Fig. 5 helps us to understand whether specific person/number combinations are responsible for these findings. With regard to (Observation 4), we see that the finding is restricted to 2sg: The comparison between 1sg (bar 1&3) and 2sg (bar 5&7) leads to a refined picture, unveiling a remarkable property of topics in contexts of domainswitching to the 2sg.

(Observation 4 rev.): In contexts of domain-switching to the 2sg, shift topics have a *lower* pronoun rate than familiar topics.

3.3.1. 2sg shift topics in domain-switching contexts

The discourse extract (17a)-(17e) shows in (17e) an example of a 2sg null subject shift topic in a domain-switching context. The example illustrates that 2sg topic referents are generally highly accessible. Neither topic shift (from speaker A in (17a) to speaker B in (17e)) nor domain-switching (the topic in (17b) to (17d) belongs to a chain of the nonlocal domain) lowers the accessibility to the point that makes the use of an overt pronoun in (17e) necessary. This would also hold if we consider the noun in the exclamative (17b) as non-referential and thus non-topical. Furthermore, ease of disambiguation does not seem to be a relevant factor in the local domain (where verbal morphology already distinguishes clearly between 1sg and 2sg).

Example (18k) in the second discourse extract (18a)–(18k) illustrates the general accessibility of 2sg referents even more clearly. (18a)–(18j) are all topicless sentences in which speaker A is asking a series of questions regarding the interior of the house that use the nonspecific subject *se* 'one' (including in the question/answer fragments in which the nonspecific subject is elided). The previous categorical sentence containing a topic, in that case a nonlocal one, occurs three utterances prior to (18a) (for reasons of space, we only reproduce (18a)–(18k)).

The last mentioning of a referent of the local domain (not even as a topic, but as part of the comment) occurs 22 utterances prior to (18a). Therefore, (18k) is preceded by a very large block of topicless sentences. If referential distance within a chain played a substantial role for the accessibility of the local 2sg referent, we would expect the use of an overt subject pronoun in (18k). In sum, we cannot identify a factor in domain-switching contexts that could explain why 2sg familiar topics have a low(er) rate. Therefore, we turn to the question as to why 2sg familiar topics have a high(er) pronoun rate in same-domain contexts.

(17a)	A:	sí, [_{Top} pro] _a	me	he	enterado	que	se	ha	muerto	[un	vecino]p
		yes pro _a	REFL	have.1SG	found.out	that	REFL	has	died	[a	néighbor] _p
		'Yes, I found o	out that a	neighbor fr	om the build	ling ha	s died.				
		del edifi	cio.						and the second s		
		from.the build	ding						·		
(17b)	B:	sí, ¡pobre [·	_{Fop} homb	re] _p !							
		'Yes, poor mar	1.1.p 1.1.'								
(17c)	A:	¿[Top pro]p era		in hombre	mayor, o?						
		'Was he an eld	erly mar	i, or?'	clucity of						
(17d)	B:	sí bueno, ([_{Top} pro]	_p) no tan	mayor, 5	6 por	ahí	tend	lría	[1	_{fop} el hombre] _p .
		yes well	prop	not_so	elderly 5	6 for	there	have	e.3sg.con	JD [t	he man] _p
		'Yes, well, not	so old, t	he man mus	t be about 50	5.'					
(17e)	A:	¿[_{Top} pro] _b lo	cono	ocías de	hace mu	cho?					
		pro _b his	n _p knev	w.28G from	n ago mu	ch					
		'Have you kno	wn him	for a long tii	ne?' (39/21)						

3.3.2. 2sg familiar topics in domain-switching contexts

The discourse extracts (19a)-(19g) and (20a)-(20j) show a pattern – question repetition for further elaboration – that seems to occur with an above average frequency with 2sg familiar topics in domain-switching contexts, as in (19g) and (20j). In (19a), speaker A first asks speaker B (the topic of the local domain), whether he saw Ester, then (in (19c) and (19e)) asks for details regarding the time and circumstances of Ester's arrival (the topic of the non-local domain), in order to repeat (in (19g)) once more the question from (19a) – presumably, she has doubts whether speaker B has reported all

A. Adli / Journal of Pragmatics xxx (xxxx) xxx

the known facts. Please note that the search domain for the topic in (19e) is the matrix and not the embedded clause (see section 1.2).

- (18a) A: ¿y de la cocina se va...? and from the kitchen SE goes 'And from the kitchen you get to...'
- (18b) B: sí, a la despensa. yes to the pantry 'Yes, to the pantry.'
- (18c) A: ¿...a la despensa que es donde estaba el arcón? to the pantry which is where was.3sG the freezer 'To the pantry, which is where the freezer was?'
- (18d) B: ajá, exacto. aha exact 'Aha, exactly.'
- (18e) A: muy bien, ¿y entonces para ir a las habitaciones que se va very well and then for go.INF to the rooms that SE goes 'Very well, so, in order to get to the rooms, you go starting from the dining room?'

desde el comedor? from the dining.room

- (18f) B: sí, por un pasillo. yes through a corridor 'Yes, through a corridor.'
- (18g) A: ¿a la derecha o a la izquierda? on the right or on the left 'On the right or on the left side?'
- (18h) B: a la derecha. on the right 'On the right side.'
- (18i) A: a la derecha. on the right 'On the right side.'
- (18j) B: m-m. 'M-m.'
- (18k) A: tal vez, $[_{Top} pro]_b$ no lo_p sepas. maybe pro_b not it_p know.2SG.SBJV 'Maybe you don't know that.' (17/132)

16 A. Adli / Journal of Pragmatics xxx (xxxx) xxx (19a) A: ¿sí [_{Top} pro]_b la_p viste llegar? herp saw.2sg arrive.INF yes prob 'Or DID you see her arrive?' B: sí. (19b) es/ (19c) '¿en qué A: momento de la noche [Top pro]p llegó? what moment of the night arrived.3sg? in ---- prop 'When exactly did she arrive last night?' (19d) **B**: media hora Bea_q y [los amigos de Bea_q]_r. con antes que with half hour before than Bea_q and [the friends of $Bea_q]_r$ 'Half an hour before Bea and her friends.' (19e) ;ah! entonces cuando Beaq [Top Ester]p tú_b fuiste de A а casa ya ah then when went.2sg house of Bea_a Estern already you_b to 'Ah! So, when you went to Bea's house, Esther had already arrived?' había llegado? had.3SG arrived (19f) B: m-m. 'M-m. (19g) lap viste? A: ¿pero [_{Top}\tú]_b no but you_b not herp saw.286 'But you haven't seen her?' (46/293)

CL

In the extract (20a)–(20j) we once again see a question asked by speaker A about what/who speaker B has seen. (20j) is not an identical repetition of (20a), but a fairly similar question. There are two topics of the nonlocal domain in between, one in (20c) (the female victim's guests, refered to by the pronoun *ellos* 'they'), one in (20h) (the female victim). The main insight from (19a)–(19g) and (20a)–(20j) is that 2sg familiar topics in domain-switching contexts occur more frequently with specific conversational patterns that favor overt realization of the pronominal subject topic. This explains why these cases have a higher pronoun rate than their shifting counterparts.

The discourse fragment (21a)–(21d) suggests that the factor favoring overt realization of the pronominal subject topics is not restricted to the repetition of questions, but to repetitions/elaborations in general, irrespective of the sentence modality. Speaker B asserts in (21a) that he would not like to tell her the facts (about her husband's extramarital affairs), and in (21d) speaker A asks him specifically whether he had said anything about it the previous day.

A. Adli / Journal of Pragmatics xxx (xxxx) xxx

- (20a) A: ¿o sea [Top tú]b lap viste viva a las cinco?
 or is.SBJV/IMP youb herp saw.2sG alive.F at the five 'In other words, you saw her alive at five?'
- (20b) B: claro. 'Sure.'
- (20c) A: $y [_{Top} ellos]_q$ al cabo de media hora se fueron? and they_q at.the end of half hour REFL went.3PL 'And half an hour later they left?'
- (20d) A: ¿y esta mañana a qué hora *pro*_r te_b han avisado? and this morning a what hour *pro*_r you_b have.3PL let.know 'And this morning, what time did they call you?'
- (20e) B: pues hacia las siete o así. then around the seven or so 'Well, around seven.'
- (20f) A: ¿de la mañana? of the morning 'In the morning?'
- (20g) B: claro, porque es cuando se levanta Ja... sure because is when REFI get.up.3SG the... 'Sure, because that's when the...gets up.'
- (20h) A: ¿o sea que en dos horas ý media a [Top esta chica]p or is.SBJV/IMP that in two hours and half.F ACC this girlp 'In other words, they killed this girl in two and a half hours?'

 pro_{s} se la_{p} han cargado? pro_{s} REFL her_p have.3PL killed

- (20i) B: claro. 'Sure.'
- (20j) A: ¿entonces cuando pro_b fuiste a la fiesta, $[T_{op}, t\hat{u}]_b$ viste quién estaba dentro? then when pro_b went.2SG to the party you_b saw who was inside 'So, when you went to the party, did you see who was inside?' (49/482)

A. Adli / Journal of Pragmatics xxx (xxxx) xxx

(21a)	B:	у	а	[_{Top} mí] _b ,	no	decirlep	nada,	pro_q	me _b	sabría	mal
		and	to	meb	not	tell:her _p	nothing	pro_q	me _b	taste.3SG.COND	bad
		'And	ł I w	ouldn't like	e not	to tell her	anything.	,			
							/				
(21b)		A: o	s	ea	que	[_{Top} la n	nujer] _p	lo _r sal	bía	entonces.	

- (21b) A: o sea que [_{Top} la mujer]_p lo_r sabia entonces. or is.SBJV/IMP that [the woman]_p it, knew.3SG then 'In other words, then, the woman/knew it.'
- (21c)B: por supuesto que $[_{Top} pro]_p$ lo_r sabía. for assumption that knew.3SG pro_n itr 'Of course she knew it.' (21d) nada no? dirías А: ¿у $[_{Top} t \dot{\mu}]_b$, ayer, DrOh. no lep
 - and you_b yesterday pro_b not her_p say.2SG.COND nothing no 'And you did not tell her anything yesterday, did you?' (3/361)

Coming back to (Observation 3), the finer-grained picture in Fig. 5 further corroborates this finding. We observe lower pronoun rates of shift topics in domain-switching contexts as compared to same-domain contexts for 1sg (bar 1&2), 2sg (bar 5&6), and, most visibly, for 3sg (bar 9&10) (somewhat minor distinctions also for 3pl, which I will leave aside in the discussion). With regard to familiar topics, we find, as expected, that domain-switching increases pronoun rate (bar 7&8, bar 11&12, bar 15&16). The notable exception are 1sg familiar topics (bar 3&4), which have a lower pronoun rate in domain-switching contexts, summarized in (Observation 5).

(Observation 5): 1sg familiar topics in domain-switching contexts have a *lower* pronoun rate than in same-domain contexts.

The following discourse fragments illustrate the peculiar behavior of 1sg familiar topics described in (Observation 5): (22f) is an example of a 1sg null subject in a domain-switching context, while (23b) is an example of a 1sg overt subject pronoun in a same-domain context.

3.3.3. 1sg familiar topics in domain-switching contexts

The first discourse fragment begins with topic referents of the local domain in (22a) (speaker A) and (22b) (speaker B).³ It then switches to topic referents of the nonlocal domain in (22d) and (22e) ('a young guy'), to switch back, in (22f), to speaker B, to whom the 1sg familiar null subject topic refers. Sentences (22d) and (22e) are about external referents. On a par with my findings with regard to 2sg topics above, 1sg topic referents are highly accessible. Even the referential distance between the two coreferential topics of the chain in the local domain created by the domain-switching does not decrease accessibility to the point that an overt subject should be used in (22f). We therefore have to place same-domain contexts under scrutiny and ask what increases the pronoun rate of 1sg familiar topics in these contexts? (23a) and (23b) illustrate this point.

³ (22c) is topicless, because the 2sg null subject of the matrix clause is not a topic candidate because the main speech act in the sense of Hooper and Thompson (1973) is the subordinate clause which only contains indefinite nonspecific arguments.

A. Adli / Journal of Pragmatics xxx (xxxx) xxx

- (22b) A: $[_{Top} pro]_a$ no creo, ;no? pro_a not think.1SG no 'I/don't think so, do you?'
- (22c) A: $\langle_{cy} pro_{b}$ crees que algún vecino haya ido alguna vez a and pro_{b} think.2SG that some neighbor has.SBJV gone some time to 'And do you think that some neighbor has ever gone to any of the parties?'

alguna de las fiestas? any of the parties

(22d) B: en el piso de arriba de todos vive [_{Top}un chico joven]_q con el_q que in the floor from above of all.M.PL lives [a guy young]_q with who_q 'On the floor above everyone lives a young guy with whom she gets along very well.'

prorsellevamuybien.prorREFLget.along.3SGverywell

- (22e) B: a lo mejor, $[T_{op} \acute{e}]_{q}$ si fue. to the best heq yes was 'Maybe it really was him.'
- (22f) A: $iy [r_{op} pro]_a$ no puedo hablar con iq_q ? and pro_a not can.1SG talk with him_q 'And can't I talk to him?' (51/322)

3.3.4. 1sg familiar topics in same-domain contexts

What is characteristic for the 1sg familiar topics in same-domain contexts is the question–answer pair. (23a) is a question uttered by speaker A who refers to speaker B by means of the pronominal topic expression $t\dot{u}$ 'you'. Speaker B responds with (23b), referring to himself with *yo* 'I'. In such question–answer pairs, speakers seem to use the overt form of the 1sg pronoun with above-average frequency when they directly answer a question addressed to them. (24a) and (24b) is also a discourse extract, with a 1sg familiar topic in a same-domain context in (24b), realized by a null form. Interestingly, this discourse snippet is not a question–answer pair, which seems to be the relevant factor behind (Observation 5).

- (23a) A: ¿no has hablado con [los vecinos]_p [Top tú]_b ya? not have.2SG talked with [the neighbors]_p you_b already
 'You haven't talked to the neighbors already?
- (23b) B: no, [Top yo]b no, no he hablado con nadie. no Ib no not have.1SG talked with nobody 'No, I haven't, I haven't talked to anybody.' (19/347)
- (24a) A: $[_{Top} pro]_a$ he llegado aquí ahora. pro_a have.1SG arrived here now 'I just arrived here.'
- (24b) A: y bueno, $[_{Top.}pro]_a$ te_b he encontrado. and well pro_a you.ACC_b have.1SG found 'And well, I found you.' (42/183)

Let us summarize the most important findings so far obtained, starting with the basic predictions from the beginning of section 3.2 on shift topics and domain-switching. First, I expected shift topics to have a higher pronoun rate than familiar topics. Secondly, I expected that the pronoun rate of topics in domain-switching contexts would be similar to same-domain contexts. Thirdly, we saw that topic continuity differs in domain-switching and same-domain contexts, and that topic continuity differs in the local and the nonlocal domain. Then, the analysis of different discourse segments brought to the fore the following noteworthy patterns.

(Observation 3):	Shift topics have a <i>lower</i> pronoun rate in domain-switching contexts than in same-domain contexts.
(Observation 4 rev.):	In contexts of domain-switching to the 2sg, shift topics have a <i>lower</i> pronoun rate than familiar topics.
(Observation 5):	1sg familiar topics in domain-switching contexts have a lower pronoun rate than in same-domain contexts.

3.4. A closer look: splitting up by topic subtypes

I have postulated in section 3.2 that *ease of disambiguation* is probably the factor behind Observation 3, i.e. shift topics in domain-switching contexts occur more frequently in discourse structures in which disambiguation of the pronominal topic referent is unproblematic. Discourse extract (16a)–(16f) showcased that the referent in the *comment* part (i.e. everything but the topic expression) occurred both in the target sentence (16f) and in the preceding sentence (16e), which makes disambiguation of the topic referent less challenging. What would be optimal for discourse structures, if disambiguation is to be easy and activation is at it's peak? These are discourse structures, in which the topic has just been mentioned in the comment part of the previous sentence. Therefore, I scrutinize, whether shift topics in domain-switching contexts occurred more frequently in such settings, and distinguish between different subtypes of shift topics in Fig. 6. I differentiate between cases in which the topic has just been mentioned in the comment part of the previous sentence (called Comment-to-Topic or Thetic-to-Topic shift) and cases in which the referent has not been mentioned in the immediately preceding sentence (called full shift). This analysis is only feasible with the data on 3sg topics, because

A. Adli / Journal of Pragmatics xxx (xxxx) xxx

the absolute frequencies with the other person/number combinations are too low to derive reliable relative frequency values.

Indeed, Fig. 6 illustrates that the difference in pronoun rate between the domain-switching and the same-domain context is most pronounced with the comment-to-topic subtype where it is more than four times higher than for the full shift subtype. This confirms my hypothesis that shift topic referents are often highly activated in domain-switching contexts.



Fig. 6. Subtypes of topic continuity.

3.4.1. (Comment-to-topic) shift topics in domain-switching contexts

The discourse extract (25a)–(25l) shows a case of comment-to-topic shift. In (25l), the 3sg subject pronoun is realized by a null form in a domain-switching context. It refers to the man of the sixth floor, mentioned in the comment part of the preceding sentence (25k). Therefore, the null subject in (25l) is a shift topic (the previous topic of the chain in the nonlocal domain refers to another group of persons introduced in (25c)), but the referent is already highly activated. Domain-switching is manifest from (25i) to (25k), where speaker B is topic.

(25a)	B:	[_{Top} el [the	señor Mister	Rodríguez] _p Rodríguez] _p	blanqueaba laundered.38G	el the	dinero money	del of.the	urbanismo urbanization
		'Miste	r Rodrígu	ez laundered th	ne money from hi	s job	with thes	e things.	,
		con with	esas these	cosas. things					
(25b)	A:	vale,	bueno.						

alright good 'Alright, well.'

		ARTICLE IN PRESS
22		A. Adli / Journal of Pragmatics xxx (xxxx) xxx
(25c)	A:	is the probability of the proba
(25d)	A:	 ¿([Top pro]_q) un par de semanas, un par de días? <i>pro_q</i> a few of weeks a few of days a few weeks, a few days?'
(25e)	B:	$([_{Top} pro]_q)$ un par de meses. pro_q a few of months 'a few months.'
(25f)	A:	$\begin{array}{llllllllllllllllllllllllllllllllllll$
(25g)	A:	bueno, ¿por qué? well for what 'Well, why?'
(25h)	A:	algún motivo así que túb conozcas, o?any motive such that youb know.2SG.SBJV or'Any motive you knew of?'
(25i)	B:	[Top y0]b_no sé muy bien. Ib not know.1SG very well 'I don't know for sure.'
(25j)	B:	puede ser que fuera por motivo de la venta de algún cuadro can.3SG be.INF that is.SBJV.IPFV for motive of the sale of some painting 'Maybe it's because of some important painting he sold, or' importante o important or
(25k)	A:	vale, $i_y = [T_{op} t\hat{u}]_b$, esta mañana, has visto [al señor del sexto] _r ? alright and you _b this morning have.2SG seen [ACC.the man of.the sixth] _r 'Alright, and did you see the man from the sixth floor this morning?
(251)	A:	$c c o [T_{op} pro]_r se llama?how p r o_r REFL calls'What's his name?' (06/302)$

Interestingly, example (15c) in section 3.2 has also been a case of comment-to-topic shift, yet one occurring in a samedomain context and realized by an overt pronoun. We often find different 3rd person referents in the immediate context of shift topics in same-domain contexts as in (15c), which leads to a more frequent use of overt pronouns compared to domain-switching contexts, where the arguments in question are a referent of the local domain and a 3rd person referent. The comment-to-topic cases are an interesting show case, because they come with equally activated topic referents. They allow us to isolate the effect of same-domain contexts and illustrate the effect of ease of disambiguation very clearly.

4. Double-chain model: a discussion

Research on information-structural notions has often paid much attention to the analysis of the sentence, but much less so to larger units. The chain structure throughout a discourse is a backbone to coherent and efficient common ground management. It carries its own reward to put more emphasis on the structure of spoken dialogues, which represents one of the most frequent discourse/text types for members of a speech community in every-day life. The aim of this study was a proof of concept for a double chain model for topics distinguishing between chains in the local domain capturing discourse participants and chains in the nonlocal domain capturing all other referents. Therefore, in sgs we annotated all topics with three chain-related features, [±shift], [±local], and [±switch] in dialogues recorded with 54 Spanish speakers. Based on these annotations, it was possible to extract a representation with a single chain as well as a representation with a double chain. Then, I determined a quantifiable criterion, the pronoun rate, in order to compare efficiency and scope of the chain models and the different combinations of chain-related features. In order to apply this criterion, I selected a subset of sentences, those in which the topic is both subject and pronominal. The analyses with pronoun rate built on 3,737 non-elliptical categorical sentences with pronominal subject pronouns.

A. Adli / Journal of Pragmatics xxx (xxxx) xxx

The results highlighted that both domain-switching and the type of domain (local or nonlocal) have specific effects on topic continuity. I identified conversational patterns that illustrate the role of domain-switching. Furthermore, familiar topics in domain-switching contexts do not behave on a par with shift topics. The findings can be more convincingly interpreted in a double TCh. Domain-switching lowers the effect of topic shift on pronoun rate. Based on a closer examination of discourse examples, I have identified, first, ease of disambiguation between different referents of the nonlocal domain as the relevant factor. Disambiguation is more challenging in same-domain contexts, where the number of third-person referents activated in the immediate context is often higher than in domain-switching contexts where a topic of the local domain comes into play. Ease of disambiguation is not a semantic notion, it does not have an effect on truth conditions. It must be carefully distinguished from contrast (Repp, 2016): The activated referents are neither alternatives – otherwise they would not have been annotated as topic in a first place because of their focus-semantic value – nor do they have contrastive discourse relations such as opposing or antithetic contributions to the question under discussion. Secondly, referents of the local domain are generally situationally evoked (Prince, 1981: 236), they generally show a high level of activation (Chafe, 1987: 25–36; Lambrecht, 1994: 106–110). They can be refered to with null pronouns even in contexts in which the distance to their last mentioning is large. 2sg familiar topics in domain-switching contexts occur more frequently in conversational contexts of repetition/elaboration, where the pronoun rate is higher. This phenomenon with familiar topics in domain-switching contexts would also not have been identifiable in a single-chain model, where they would be annotated on a par with shift topics. With regard to 1sg familiar topics in same-domain contexts, I observed that they occur more frequently with question-answer pairs, which come with a higher pronoun rate.

Two of the three highlighted findings that deviate from the basic predictions, (the revised version of Observation 4 and Observation 5), refer only to TChs in the local domain. A double-chain model is a prerequisite for bringing these differences between domains to the fore. We reach the conclusion that speaker and hearer operate in their common ground management with two parallel or superposing prominent referential structures, one linked to the speakers themselves and one to the entities they talk about. This view also suggests that the interaction between the two information-structural notions topic and givenness deserves further attention, because the entities of the local domain have a special status with regard to givenness. Future research needs to assess the role of distance between the referents in a topic chain, in particular for the nonlocal domain. On a more general level, it suggests that theories of discourse structure and coherence (e.g. Centering Theory) benefit from treating referents in the local and nonlocal domain differently. One of the open questions for future research is whether situationally evoked/given 3rd person referents behave similarly to topic referents of the local or similarly to topic referents of the nonlocal domain. We conclude that the double-chain model seems to be a promising path towards a more fine-grained analyses of dialogues.

Acknowledgements

I want to thank my student assistant Lena Pfannholzer. Without her help and patience, the visual representation of the referential chains would not look so pretty. I also want to thank Eric Engel for his precise feedback on the topic algorithm, to conform with the flowchart ISO. This research has been funded by the German Research Foundation.

References

Adli, Aria, 2011. Gradient Acceptability and Frequency Effects in Information Structure: A Quantitative Study on Spanish, Catalan, and Persian. Habilitationsschrift, Universität Freiburg.

- Brunetti, Lisa, 2009. On the semantic and contextual factors that determine topic selection in Italian and Spanish Linguist. Rev. 26 (2–3), 261–289.
- Chafe, Wallace L., 1976. Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In: Li, C.N. (Ed.), Subject and Topic. Academic Press, Inc., New York, San Francisco & London, pp. 25–55.
- Chafe, Wallace L., 1987. Cognitive constraints on information flow. In: Tomlin, R.S. (Ed.), Coherence and Grounding in Discourse. John Benjamins, Amsterdam, pp. 21–52.
- Diessel, Holger, 2012. Deixis and demonstratives. In: Maienborn, C., von Heusinger, K., Portner, P. (Eds.), Semantics: an International Handbook of Natural Language Meaning, vol. 3. Mouton de Gruyter, Berlin, pp. 2407–2432.

Frascarelli, Mara, 2007. Subjects, topics and the interpretation of referential *pro*: an interface approach to the linking of (null) pronouns. Nat. Lang. Linguist. Theory 25 (4), 691–734.

Grosz, Barbara J., Joshi, Aravind K., Weinstein, Scott, 1995. Centering: a framework for modeling the local coherence of discourse. Comput. Linguist. 21 (2), 203–225.

Harley, Heidi, Ritter, Elizabeth, 2002. Person and number in pronouns: a feature-geometric analysis. Language 78, 482-526.

Hoffman, Beryl, 1998. Word order, information structure, and centering in Turkish. In: Walker, M.A., Joshi, A.K., Prince, E.F. (Eds.), Centering Theory in Discourse. Clarendon Press, Oxford, pp. 251–272.

Holmberg, Anders, 2005. Is there a little pro? Evidence from Finnish. Linguist. Inq. 36 (4), 533-564.

Holmberg, Anders, Nayudu, Aarti, Sheehan, Michelle, 2009. Three partial null-subject languages: a comparison of Brazilian Portuguese, Finnish and Marathi. Stud. Linguist. 63 (1), 59–97.

Hooper, Joan B., Thompson, Sandra A., 1973. On the applicability of root transformations. Linguist. Inq. 4 (4), 465-497.

Jacobs, Joachim, 2001. The dimensions of topic-comment. Linguistics 39 (4), 641-681.

Krifka, Manfred, 2014. Embedding illocutionary acts. In: Roeper, T., Speas, M. (Eds.), Recursion: Complexity in Cognition. Springer International Publishing, pp. 59–87.

Labov, William, 1972. Sociolinguistic Patterns. University of Pennsylvania Press, Philadelphia.

Bianchi, Valentina, Frascarelli, Mara, 2010. Is topic a root phenomenon? Iberia. Int. J. Theor. Ling. 2 (1), 43-88.

Krifka, Manfred, 2007. Basic notions of information structure. In: Féry, C., Fanselow, G., Krifka, M. (Eds.), Working Papers of the SFB632, Interdisciplinary Studies on Information Structure (ISIS) 6. Universitätsverlag Potsdam, Potsdam, pp. 13–56.

A. Adli / Journal of Pragmatics xxx (xxxx) xxx

Lambrecht, Knud, 1994. Information Structure and Sentence Form. Topic, Focus and the Mental Representations of Discourse Referents. Cambridge University Press, Cambridge.

Otheguy, Ricardo, Zentella, Ana Celia, Livert, David, 2007. Language and dialect contact in Spanish in New York: toward the formation of a speech community. Language 83 (4), 770–802.

Prince, Ellen F., 1981. Toward a taxonomy of given-new information. In: Cole, P. (Ed.), Radical Pragmatics. Academic Press, New York, pp. 223-255.

Reinhart, Tanya, 1981. Pragmatics and linguistics: an analysis of sentence topics. Philosophica 27, 53–94.

Repp, Sophie, 2016. Contrast: dissecting an elusive information-structural notion and its role in grammar. In: Féry, C., Ishihara, S. (Eds.), The Oxford Handbook of Information Structure. Oxford University Press, Oxford, pp. 270–289.

Ritter, Elizabeth, Wiltschko, Martina, 2009. Varieties of INFL: TENSE, LOCATION, and PERSON. In: Craenenbroeck, J.v. (Ed.), Alternatives to Cartography. Mouton de Gruyter, Berlin, pp. 153–201.

Roberts, Craige, 2011. Topics. In: Maienborn, C., Heusinger, K.v., Portner, P. (Eds.), Semantics: an International Handbook of Natural Language Meaning, vol. 33/2. Mouton de Gruyter, Berlin, pp. 1908–1934.

Sasse, Hans-Jürgen, 2006. Theticity. In: Bernini, G., Schwartz, M.L. (Eds.), Pragmatic Organization of Discourse in the Languages of Europe. Mouton de Gruyter, Berlin & New York, pp. 255–308.