Antitopic and Word Order in Conversational French

Introduction

In their 1976 study "Subject and topic: A new typology of language," Li and Thompson (1976) classify languages based on the prominence of topic-comment constructions versus subject-predicate constructions, and provide eight criteria for determining whether a language is topic-prominent, subject-prominent, or one of two intermediate stages. It may be possible to convert these criteria into quantitative tests for topic-prominence. In this study I propose one possible set of quantitative tests and apply it to conversational French, a language which some have speculated may be in the process of moving from the subject-prominent type to the topic-prominent. The study reveals that the emergence of antitopic constructions in conversational French may be connected with the lack of change in these criteria.

Background

Subject and Topic

Li and Thompson's typology is based on the organization of information, distinguishing the two categories of topic-prominent (e.g. Mandarin) and subject-prominent (e.g. English) languages. These are supplemented by the intermediate cases of both topic and subject-prominent (e.g. Korean) and neither topic nor subject-prominent (e.g. Philippine languages). Li and Thompson further hypothesize a life-cycle of information organization, in which topic-comment constructions emerge from discourse and are then grammaticized into subject-predicate constructions. After this grammaticization, there is usually independent motivation for a new generation of topic constructions to emerge from discourse as the subject-predicate constructions grammaticize into agreement marking.

Despite being listed among the extreme subject-prominent languages by Li and Thompson (page 483), French is in fact a possible example of this cycle in action. Although Classical French is thought to have been a strict subject-prominent language, there is evidence that contemporary spoken French may be changing towards a structure where both topic-comment and subject-predicate constructions are prominent, and eventually towards pure topic-prominence. This study aims to investigate this possibility.

There is a lot of confusion surrounding the semantics of topic-comment constructions, and whether all of the constructions that have been labeled as "topics" are semantically equivalent. For the purpose of this study, I will lay the semantic considerations aside and use a more syntactic definition of subject. I will therefore consider as a topic any noun phrase that does not have a direct (i.e. subject or object) syntactic relation to the verb, even though it may be co-referential with the subject or object of the clause.

Spoken French vs. Written French

Although spoken French has been continuously changing at a relatively normal pace, written French has changed comparatively little since it was standardized centuries ago. The gap between spoken and written forms was recognized in 1921 by Vendryes, who wrote (page 171, translation mine), "In French, written and spoken language are so far removed from one another that people never speak the way they write and rarely write the way they speak." Such basic grammatical features as negation (*pas* vs. *ne...pas* for predicate negation), interrogation (intonation vs. inversion for yes-no questions) and the pronominal system (*on* vs. *nous* for first-person plural) have significantly different forms in the spoken language and the written language. In fact, a second set of writing

conventions is emerging in France, as more popular written forms such as comic books and advertisements are written in a form that more closely approaches the spoken language.

The topic-comment constructions mentioned above are attested by Vendryes in 1921, and are becoming more common in the spoken language, but not in the written language. Written French still requires a rigid subject-predicate structure for almost all sentences, but it occasionally employs noun phrase topics to set up a theme for an extended section of text. By contrast, these topics are more frequently heard in spoken French, where they can serve to narrow the scope of a single sentence. Topics in French almost always are followed by an overt or pronominal subject and marked with a conventionalized intonation contour. They always occur at the beginning of a clause, and are often co-referential with one of the verb arguments. Their increased frequency in spoken French is often linked to the grammaticization of pronominal clitics into preverbal agreement markers; as subject pronouns become more obligatory and less distinctive, topics arise to take the place of subjects.

Topic vs. Antitopic

Another phenomenon occurring in contemporary spoken French that has received relatively little attention is postposed noun phrases, called "antitopics" in Chafe's (1976) discussion of Seneca, and observed in French by Lambrecht (1981). Similar to topics, they are often co-referential with one of the verb arguments, but they usually appear at the end of a clause. I will not address their semantic and pragmatic functions, other than to say that they appear to be similar to the functions of topics. Antitopics were also

observed in 1921 by Vendryes, who borrows from Bally (1913) an example of the spoken

French of that period:

Son   enfant! Mais elle   le déteste, cette mère.
3sg.POSS child   but 3sg DET hate   that mother
"But that's a mother who really hates her kid."

        We will see in this study that the similarity of function between topics and

antitopics may have played a significant role in recent changes in French syntax.

Methods

A Quantitative Application of Li and Thompson's Theory

        Li and Thompson propose what are essentially two continua between topic-

prominent and subject-prominent languages, and argue that languages move along one

continuum from topic-prominent to subject-prominent via the "neither" stage, and along

the other from subject-prominent to topic-prominent through the "both" stage. Since they

identify the relative position of various languages along this continuum, it should be

possible to more specifically measure this position, in fact contrasting it over time.

        On pages 466-471, Li and Thompson outline a number of factors distinguishing

topic-prominent languages from subject-prominent languages, as follows:

a) Surface coding of the categories topic or subject. Topic-prominent languages have a specific coding for topics, and subject-prominent languages for subjects.
b) The use of the passive construction. Topic-prominent languages are able to use the passive less because this information can be conveyed by placing the patient in topic position.
c) "Dummy" subjects. Subject-prominent languages use more dummy subjects because subjects are more obligatory.
d) "Double subject" constructions. Topic-prominent languages often have full noun phrases in both topic and subject position.
e) Controlling co-reference. "In a topic-prominent language, the topic, and not the subject, typically controls co-referential constituent deletion" (page 469).
f) Verb-final word order. Topic-prominent languages tend to have verb-final as their unmarked word order.

g) Constraints on the topic constituent. Subject-prominent languages tend to have more constraints on what can appear in the topic position.
h) Basicness of topic-comment sentences. In topic-prominent languages it is more straightforward to say that subject-predicate structure is derived from topic-comment structure than the other way around.

In their discussion of transitivity and discourse, Hopper and Thompson (1980) not only list a number of characteristics of transitive constructions, but analyze the frequency of occurrence of those characteristics. What follows is an attempt to adapt Li and Thompson's factors of topic-prominence in a similar quantitative fashion. Of these factors, it is possible to test for five of them in a corpus. They can be measured by the frequency of occurrence of a construction as a proportion of the number of times such a construction is syntactically possible. This study will test each of these factors on a corpus of French discourse, with a different means of determining the number of possible occurrences. The tests for these factors are as follows:

1) Surface coding. Whether constructions exist that use topic or subject positions. This is a yes-no question, not a quantitative measure.
2) Basicness of the topic-comment construction. The token frequency of topic constructions, as a proportion of the total number of clauses in the text.
3) Passive constructions. The frequency of passive constructions as a proportion of the number of times a passive or transitive construction occurs in the text.
4) "Dummy" subjects. The frequency of dummy subjects as a proportion of the number of impersonal constructions.
5) Verb-final word order. The frequency of clauses where no object follows the verb, as a proportion of the total number of transitive constructions.

It should be possible to devise tests for the remaining factors, "double subject" constructions, controlling co-reference and constraints on the topic constituent, but there were unfortunately not enough topic-comment constructions in the corpus, with the possible exception of the data taken from François 1974 (20 tokens) to allow a meaningful analysis.

An additional factor was also measured:

6) Antitopic constructions. The similarity in form and function between these constructions and the topic constructions suggests a possible relationship, so these were counted as well, as a proportion of the total number of clauses in the text.

Data Sources

Since this phenomenon has been observed mostly in conversational French, this study will examine data from several sources of conversation. Samples from the conversation of two present-day French speakers will be analyzed. Conversational discourse from 1964 will be compared with the present-day data to determine the extent to which the language has changed over the past thirty-five years. A contemporary non-conversational text will also be compared with the present-day conversational samples to determine the effect of style.

François 1974

This corpus contains phonetic transcripts of three casual conversations among members of François' extended family in 1964. The conversation involves relaxed discussion about their hometown of Argenteuil in the Paris suburbs, its history, a younger family member's performance at school, and gardening. Because of difficulties comparing or combining samples from two participants in the same conversation, I have chosen to focus exclusively on the speech of one of the participants, LS, a sixty-five-year-old man, taken from conversations with his wife and François herself.

Internet Relay Chat

As a convenient source of present-day conversational data, I have chosen to examine logs of Internet Relay Chat (IRC) discussions. IRC is a protocol that allows computer users from around the world to communicate real-time via the Internet by typing messages to each other. There are often several participants on a channel at a

given time, and often several conversations take place simultaneously. Unlike on-line

discussion forums such as Usenet, the real-time nature of IRC encourages a more

conversational style among its participants. The conversations studied were all taken

from the IRC channel #france, a general-interest French-language forum.

As is typical for IRC, the members of #france have developed a style of language

that differs from the conventional written form. Word order and vocabulary closely

mirror the spoken language. A number of abbreviations are used for frequent words and

phrases, such as "c" for "c'est" (it is), and "a+" for "à plus tard" (see you later). English

is a clear presence: English speakers often join the channel asking for French lessons, or

advertising pornography. Several conventions are borrowed from English-language IRC,

such as "lol" for "laughing out loud" and smiley faces such as ":~(." Participants will

occasionally code-switch to English, especially to quote an American song or other text.

This study will examine the conversations of two members of #france: Yield_, a

21-year-old male heavy metal fan, according to his statements on #france, and Dr_Graph

(also known as Dr-Graph), a 22-year-old male computer technician according to his web

page (http://www.chez.com/drgraph). Both live in France, and as is common on IRC,

both chose to use a "nick" (often using non-alphabetic characters) to identify themselves

rather than giving their true names. Yield_'s data were collected from the evenings of

April 12-13, 1999, and Dr_Graph's from the evening of April 25, 1999. Although

Dr_Graph did briefly join #france on April 12, he had no conversation with Yield_, so

their conversations can be considered independent.

The following is an excerpt from the log for April 12, 1999; it contains part of a

discussion between brice and Dom-- about Dom--'s severe toothache, and a discussion

between Yield_ and Babook about hard rock music ("SPKR" indicates the name of the

speaker, "ADREE" the name of the addressee):

<brice> dom-- alors va a l' hostau ca   va faire   plaisir aux medecins des urgences)
SPKR ADREE so   go to DET hospital that do.3sg.FUTP pleasure to.DEF.PL EMS
"So go to the hospital; that will make the EMS happy."

<Dom--> non, a 1h   du   mat, c'est aps   trop   mon   genre
Speaker   no   at one.hour of.the morning it's not too.much 1sg.POSS style
"No, at one AM, it's not really my style"

<Yield_> qd   tu   joueras   sur Panam fais   moi   signe
SPKR   when you   play.FUT.2sg on NAME do   1sg.OBL sign
"When you play at Panam, wave to me."

<Yield_> c koi   le   nom de   ton   groupe ?
SPKR   it's what DET.M name of 2sg.POSS group
"What's the name of your group?"

<brice> alors   tu   es   condamne !!
SPKR   so 2sg.SUBJ BE condemn.PPT
"So you're screwed!"

<Babook> yield : ca risque pas !!!   j'ai   pas de   groupe ici !! :~(
SPKR   ADREE that risk NEG   1sg.have NEG.PART group here unhappy
"Not gonna happen! I don't have a group here! :~("

        At this point, IRC French can not be said to be equivalent to spoken

conversational French. For the purposes of this pilot study I am assuming that the two

varieties are comparable, but further work needs to be done to show the extent to which

IRC data may be used interchangeably with spoken data.

Le Monde des Livres
        For stylistic comparison, an example of contemporary normative written French

was taken from the periodical *Le Monde des Livres* for April 30, 1999. "La Société de

Bataille" was a discussion by Phillippe Sollers of the role and legacy of the author

Georges Bataille, and was obtained from the web site of Le Monde,

<http://www.lemonde.fr>.

Data Analysis

Conversational Style: Yield_ vs. Dr_Graph

My first objective is to determine whether any of the characteristics of the IRC

participants' language use can be attributed to individual style, rather than to an overall

trend in IRC French. I therefore compared the contributions of Yield_ with those of

Dr_Graph by the measures indicated above. It seems appropriate to compare these two,

because although they are two of the most prolific contributors to #france, intuitively

Dr_Graph's style gives the impression of being more formal, more careful and closer to

written-language norms. The following table compares the two participants, using the

Chi-square Test of Association ($\alpha = 0.05$).

Factor	Yield_		Dr_Graph		Significance test	
	Percent	n	Percent	n	χ^2	p
Surface coding	Topic and Subject		Topic and Subject		N/A	N/A
Topic constructions	3.85%	4	3.83%	7	<0.001	p>0.50
Passives	2.22%	1	3.33%	3	0.13	p>0.50
Dummy subjects	0.00%	0	40.00%	4	N/A	N/A
Verb-final order	22.73%	34	29.89%	61	0.75	0.50>p>0.25
Antitopics	7.69%	8	4.37%	8	1.39	0.25>p>0.10

The difference between the two participants fails to achieve statistical significance

for any of the measures. In the case of dummy subjects, the statistical test is inapplicable,

since none of the impersonal verbs were produced with dummy subjects by Yield_.

From this point on I will use Yield_'s data as representative of the usage on #france.

Effect of the Norm: Yield_ vs. Sollers

Since the gap between formal written and conversational spoken French has been well-documented, I compared Yield_'s conversation to Sollers' article in Le Monde, a daily newspaper with a reputation for conservative writing. The following table summarizes the results ($\alpha = 0.05$).

Factor	Yield_		Sollers		Significance test	
	Percent	n	Percent	n	χ^2	p
Surface coding	Topic and Subject		Topic and Subject		N/A	N/A
Topic constructions	3.85%	4	1.96%	2	0.65	p>0.25
Passives	2.22%	1	25.00%	10	9.75	0.01>p>.001
Dummy subjects	0.00%	0	87.50%	7	N/A	N/A
Verb-final order	22.73%	34	63.33%	19	1.71	0.25>p>0.10
Antitopics	7.69%	8	0.00%	0	N/A	N/A

Passive constructions are almost nonexistent in Yield_'s conversation, but constitute a full quarter of the forty constructions in Sollers' piece where the construction is possible. It is important to note here that contrary to the teachings of American rhetoric, passives are fully acceptable in French writing, and judicious use of the passive is considered a mark of skillful writing.

As in the earlier comparison, the fact that Yield_ uses no dummy subjects whatsoever prevents us from using a chi-square test to find the significance of the difference between Yield_'s conversation and Sollers' article. Sollers does use dummy subjects in seven out of the eight impersonal constructions in his article, the one exception being *voilà*, "there is," which is derived from an imperative and is forbidden from taking a subject by the norms of written French. Discounting *voilà*, we find all dummy subjects in Sollers' writings, and none in Yield_'s.

The difference in use of topics and of verb-last constructions is not significant between the two styles. As we will see later, this provides a challenge for the connection between topic constructions and word order.

A Diachronic Test: Yield_ vs. LS

In order to determine what changes may have occurred in these parameters over the past thirty-five years, I compared the sample of Yield_'s conversations with data from LS's conversations in three-sevenths of the sections in François' Corpus I. The following table shows the differences between the two ($\alpha = 0.05$).

Factor	Yield_		LS		Significance test	
	Percent	n	Percent	n	χ^2	p
Surface coding	Topic and Subject		Topic and Subject		N/A	N/A
Topic constructions	3.85%	4	7.94%	20	1.96	0.25>p>0.10
Passives	2.22%	1	7.87%	10	1.77	0.25>p>0.10
Dummy subjects	0.00%	0	20.00%	3	N/A	N/A
Verb-final order	22.73%	34	55.56%	52	13.85	0.001>p
Antitopics	7.69%	8	0.40%	1	15.90	0.001>p

The number of topic constructions and passives are not significantly different, and the number of dummy subjects used by LS is not very large even by comparison to LS's use of none at all. The main factor that has changed significantly since 1964 is in fact the use of verb-final word order, which is lower in Yield_'s data than in LS's. The use of antitopics by Yield_ is also significantly higher than the single one used by LS; in fact, the antitopic used by LS may in fact be an appositive:

Y avait   sa   fille,   une grande (incomprehensible).
EXIST 3sg.POSS daughter DET tall   Ö
"Then there was his daughter, a tall (incomprehensible)."

Summary

The data examined show that with respect to the factors identified by Li and Thompson as related to topic-prominence, the difference between two conversational written (IRC) French writers is not significant, but the difference between the IRC register and normative written French is significant for at least two of Li and Thompson's factors, dummy subjects and passives. This confirms the widespread impression of a gap between conversational and standard French relating to topic constructions.

The diachronic data show an interesting pattern, however. Topic-comment structures have not increased in frequency at all over the past thirty-five years; in fact, it is not impossible that they could have decreased in frequency. What has increased significantly in frequency since 1964 is the frequency of antitopic constructions. One possible explanation for this change is that antitopics fulfill a function very similar to that of topic constructions in conversational French. Once the antitopic construction became well-established, it may have become the preferred alternative to topic constructions.

It is also interesting that verb-final clauses have not increased in frequency significantly since 1964. This fits with Li and Thompson's explanation for the association between verb-final order and topic-prominence (page 484):

> … in propelling a language from stage (C) [subject-prominence] through stage (D) [both] and then to stage (A) [topic-prominence], the sentence type that plays a major role is the "double subject" type of sentence. The more such sentences are used in the language, the closer the language comes to stage (A), since these are topic-comment structures par excellence. Now note that the "double subject" constructions are always of the form [$_{\text{topic}}$ NP$_1$] [$_{\text{comment}}$ NP$_2$ V], which is precisely the typical sentence structure of a verb-final language. This sentence type becomes pervasive as the relationship between NP$_1$ and NP$_2$ becomes less and less constrained.

If Li and Thompson's analysis is correct, the increased use of antitopic constructions, which come at the end of a clause, would tend to push the language away from verb-final word order and towards verb-medial, if not verb-initial, word order. This also explains why there is no significant difference in word order between Yield_'s conversation and Sollers' newspaper article.

Future Considerations

This study in its current form is essentially a pilot study. This section will therefore outline a number of improvements that can be made to the study in the future, as well as directions for future research.

A Stronger Data Base

One of the most tenuous assumptions in this study is the equivalence of written conversational French (IRC) with spoken conversational French. Use of contemporary spoken French would resolve this issue. Spoken French data could be used to replace the IRC data, or because IRC data is easier to collect, enough spoken French data could be collected to provide a conclusive test of the equivalence of the two genres for these measurements.

Whether spoken or written, a larger database of conversational French would provide more support for the conclusions here. If it is possible to find other conversational data from the same period as François' corpus, this would enable us to compare sample means, rather than single speakers.

Whatever the data source, a larger number of tokens of topic-comment constructions would enable us to test for Li and Thompson's other three criteria (double subjects, controlling co-reference and constraints on the topic constituent), which would

provide a more complete measure, especially in the later stages of Li and Thompson's hypothesized change.

Since the data used in this study show no change in the token frequency of topic-comment structures, we can speculate that there was an increase some time in the past few hundred years. Transcripts of earlier conversations, if available, might help shed some light on these earlier changes.

Tests for the Measures of Topic-Prominence

The measurements that were used to test topic-prominence in this study are based on criteria outlined on pages 466-471 of Li and Thompson. To test the reliability of these measures, it would be useful to apply them to the sample of languages that Li and Thompson based their criteria on. Specifically, applying these measures to extreme topic-prominent (such as Lisu or Mandarin) and subject-prominent languages (such as English or Twi) would allow us to establish endpoints for this change, and thus get a more accurate picture of the continuum.

Antitopic is another phenomenon worth investigating in greater depth. The term was invented by Chafe (1976) to describe the Seneca language, and then adapted by Lambrecht to describe spoken French. It would therefore be useful to compare French with other antitopic languages like Seneca, and thus gain a greater insight into the workings of antitopic languages.

Conclusion

The results from this study show that the set of quantitative measurements developed above may well be an accurate measure of topic-prominence in language. If these measurements are reliable, then French has not become significantly more topic-

prominent in the past thirty-five years. This may be related to the increase in use of

antitopic constructions over the same time period, which may have fulfilled similar

pragmatic functions to the topic-comment constructions while blocking, instead of

encouraging, a change to verb-final word order.

References

BALLY, CHARLES. 1913. Le langage et la vie. Geneva.

CHAFE, WALLACE L. 1976. Givenness, contrastiveness, definiteness, subjects, topics and
points of view. In Charles Li, ed. Subject and topic. New York: Academic Press.

FRANÇOIS, DENISE. 1974. Français parlé. Paris: SELAF.

HOPPER, PAUL AND SANDRA THOMPSON. 1980. Transitivity in grammar and discourse.
Language 56, 251-299.

LAMBRECHT, KNUD. 1981. Topic, antitopic and verb agreement in non-standard French.
Amsterdam: John Benjamins.

LI, CHARLES AND SANDRA THOMPSON. 1976. Subject and topic: A new typology of
Language. In Charles Li, ed. Subject and topic. New York: Academic Press.

VENDRYES, J. 1921. Le langage: Introduction linguistique à l'histoire. Paris: La
Renaissance du Livre.