THE EMERGENT SYNTAX
OF PRE-SENTENTIAL TURN OPENINGS

Anne VICHEP and David SANKOFF*

Received February 1987; revised version February 1988

Spoken French contains a rich variety of pre-sentential discourse particles playing structural,
interactional and pragmatic roles. In 65% of some 3700 utterances in classroom discussions and
ordinary conversations, of both native and non-native French speakers, we find intonationally
unitary turn-openings containing up to five of these particles. Distributional analysis reveals
strong ordering constraints among the particles, with eight paradigmatic 'slots' available. The first
two positions structure the transition from hearer's to speaker's roles (kiaaia.... ah bon...).
Usually, the following two positions allow a bi-morphemic reaction (mais oui.... ben non...) to the
interactional and illocutionary validity of the preceding turn, while the fifth and sixth slots contain
analogous reactions to its informative content. The remaining slots contain tag-like elements
which play a largely prosodic role. Quantitative analysis of particle co-occurrence is based on the
null hypothesis that occurrences in different slots are statistically independent. A comparison of
observed versus predicted frequencies detects particle combinations used more, or less, than could
be accounted for by the frequencies of their individual components, suggesting emergent
colleations, and functionally incompatible components, respectively. Finally, we give a brief
analysis of the semantic-pragmatic functions of some particle combinations.

1. Introduction

A great variety of words and expressions are incorporated into turn openings
in spoken French. Some of these are illustrated by the segments in capital
letters in the conversational fragment consisting of examples 1-15. 'Frozen'
expressions containing more than one word are underlined.

(1) BON BEN ALORS moi tout-de-suite / j'vais j'vais entamer comme ça
euh...
'VERY WELL THEN' me right away / I am going I am going to start like this uh...

(2) OUI MAIS ça dépend / d'abord euh i faut savoir c'que c'est que la libération d'la femme //
'YES BUT it depends / first uh you must know what it is, women's lib /'

(3) NON MAIS D'TOUT MANIERE la libération d'la femme elle s'ra acquise seulement quand l'homme il aura pris conscience de ...
'NO BUT ANYWAY women's lib will be achieved only when men have become conscious of...'

(4) MAIS OUI MAIS C'EST SSS
'BUT YES BUT (OF COURSE)'

(5) BEN OUAIS MAIS C'EST SUR / c'est une prise de conscience...
'WELL YES BUT OF COURSE / it's arriving at a consciousness....'

(6) OH BEN OUI HEIN i n'changeront pas / i s'woudront toujours des hommes /
'OH WELL YES EH they won't change / they'll always want to be men /'

(7) OUAIS / ça l'apprend à la vie d'moins tard...
'YES , it'll teach him for later on in life...'

(8) AH BEN OUI MAIS / MAIS JU TEMENT j'lui filerai une poupée — si ça peut lui faire plaisir...
'AH WELL YES BUT / BUT EXACTLY I'll give him a doll — if that makes him happy...'

(9) MAIS D'ACCORD / MAIS / MAIS OUI MAIS / au fond d'toi-même tu t'diras / ah ben c'est bizarre j'ai un garçon qui veut une poupée euh c'est pas normal /
'BUT OF COURSE / BUT / BUT YES BUT / deep down you yourself would say / ah well it's weird I have a boy who wants a doll uh it isn't normal /

(10) OUI OUI MAIS TU SAI HEIN / c'qui est normal et pas normal c'est pas évident hein / / pasque...

(11) OUAIS
'OH YES BUT YOU KNOW EH / what's normal and not normal is not obvious eh/ / because...'

(12) ET PI t'as l'école aussi...
'AND THEN you have the school as well...'

(13) EH BEN TU VOIS c'est pas d'la tarte ces problèmes là /
'OH WELL YOU SEE it's not a piece of cake these problems /

(14) Ô i aura toujours des gens qui n'évolueront pas /qui s'ront contre /
'There will always be people who will not change / who will be against /

(15) NON MAIS ATTENDS / (...) (...) 'NO BUT WAIT /

[following turn]
In this paper we report on an extensive study of such turn openings based on a corpus of recorded and transcribed spontaneous dinner-table conversations and more focused discussions on specific themes such as women’s liberation, prospects for the future, etc. The participants in the conversations were largely middle class, while the discussions involved high school and university students from various social backgrounds. Ages ranged from 15 to 65. About three quarters of the material is from native speakers of French and the rest from fluent second language French speakers.

2. The forms

Examining the component parts of the turn openings, we are able to count from 50 to 80 different individual forms, depending on how we collapse similar variants and distinguish between functionally different homonyms. We have provided largely etymological glosses for the forms in the examples, although these can sometimes be functionally and semantically misleading. Bon as in (1) is literally ‘good’ but is used like ‘OK’ or ‘all right’. Ben in (1), (5), (6), (8) and (13) is similar to ‘well’, as studied by Schiffrin (1985), but is distributed differently. Oh as in (6), (10) and (11) can be translated by its English homograph, though in Schiffrin’s (1984) terms, the management of information that makes use of this form does not always coincide in the two languages. Justement means ‘exactly’ but is often used not in agreement, but to turn the interlocutor’s own argument against him or her (Cadiot, Ducrot, Récanati and Vicher (1982)). Tu sais (10) ‘you know’ and tu vois (13) ‘you see’ have similar usage in English (Erman (1986), Östman (1981)). Hein as in (6) and (10) more or less corresponds to Canadian ‘eh’, and functions somewhat like phrase-final intonation patterns reported for Australian and other dialects of English (Guy, Horvath, Vonwiller, Daisley and Rogers (1986)). Mais, analysed in detail by Anscombe and Ducrot (1977) and Ducrot (1980), has much the same effect as ‘but’ as described by Altenberg (1986), Baker (1975) and Quirk (1955); de toute façon and de toute manière correspond, in some contexts, to ‘anyway’ (Altenberg (1986), Quirk, Greenbaum, Leech and Svartvik (1972)).

Under the rubric of turn openings we include not only those utterances which successfully initiate full turns but also those like (15) which fail, as well as utterances like (11) which are not real attempts to take the floor but rather fall into the category of what Tannen (1984) and others call back-channel activity. We agree with Sacks, Schegloff and Jefferson (1974) that these reduced forms, which occur mostly at breaking or transition points, should be considered as turns. It is an empirical question whether they tend to be realized by different forms.
One characteristic of turn openings in (European) French, at least quantitatively different from other speech communities which have been studied, is the great tendency to combine two, three, or more forms in a single turn opening. This is true of all these examples shown above, except (7) and (11), which contain single terms, and (14), where there is no marked turn opening.

3. Distributional analysis

Our main goal in this study is to characterize the processes which result in these combinations, and to see how the semantic, argumentative and interactive functions of the turn opening relate to the syntagmatic organization of the constituent forms. Our methodological strategy was to proceed as far as possible with a purely distributional analysis of surface form, then to try to extract the major combinatorial tendencies using specially devised statistical tools, and finally to interpret the results within the framework of Anscombre and Ducrot's 'integrated pragmatic' theory of \textit{énonciation} \footnote{Énonciation focuses on the act of producing an utterance as well as the attitude of the speaker towards this utterance. Thus we will distinguish between the propositional content of an utterance and its 'enunciative framework', the argumentative aspects of which Lyons (1977) calls the 'act of utterance', in his exposition of this theory.} (Ducrot (1980), Anscombre and Ducrot (1983), Ducrot (1984)).

A first distinction to be made is between single-utterance combinations, possibly composed of several terms, and series of two or more turn openings clearly separated prosodically. This is illustrated in numbers (8) and (9). In (8) we would consider \textit{ah ben oui mais} and \textit{mais justement} as two openings. In (9) \textit{mais d'accord, mais} and \textit{mais oui mais} are three openings. In this study we analysed each prosodically distinct utterance separately.

Examples (16)-(22) illustrate the distributional analysis, for which we shall present only the broad outlines.

(16) \textit{ah bon}
(17) \textit{bon ben}
(18) \textit{ah ben}
(19) \textit{ah bon ben}
(20) \textit{*ben bon}
(21) \textit{*bon ah}
(22) \textit{*ben ah}

\footnote{The 66 participants in our corpus were mostly French (65%); there are however some Belgians (10%) and Swiss (3%). The remaining speak French as a second language and come from the USA, Portugal, Poland, Ireland, Uruguay and the Flemish part of Belgium.}
All of the orderings in (16) to (19) were well attested, while (20), (21) and (22) did not occur at all. It follows from these examples that the relative positions of whichever two or three terms occur are constrained to be as in (19). Other forms, such as *oh* or *eh*, are found to occupy the same slot as *ah*, while *mais* can occur in the same position as *ben*, but not before *bon* or *ah*.

Since some turn openings had as many as five component terms, e.g. *ah bon ben oui tu vois*, it was necessary to postulate at least five syntagmatic slots to account for the data. We first tried to see whether all the examples could be handled within this framework, i.e. whether a given term could always be assigned to the same slot, e.g. *bon* in (23) and (25), *ben* in (23) and (24) and *ah* in (23) and (24), conserving the same order in the data utterance as in the model sentence.

(23) AH BON BEN OUI TU VOIS
(24) AH BEN NON
(25) BON D'ACCORD HEIN

This was not sufficient to accommodate all the examples, however, such that each form always fell into the same column. Indeed, it was found that elements of one series of terms, namely *mais* and *ben*, could be used in two distinct places in the utterance and similarly for another series including *oui*, *non* and *d'accord* as shown in the following:

(23) AH BON BEN OUI TU VOIS
(24) EH BEN NON
(25) BON D'ACCORD HEIN
(26) MAIS OUI MAIS D'ACCORD
(27) OUI BEN JUSTEMENT TU VOIS
(28) OH OUI MAIS TU SAIS HEIN

Finally eight slots were needed to provide an analytical grid capable of handling all the data, as illustrated by the examples in (29):

(29) (1) (2) (3) (4) (5) (6) (7) (8)
    AH BON BEN JUSTEMENT TU VOIS
    BEN OUI MAIS JUSTEMENT
    MAIS OUI MAIS D'ACCORD
    MAIS OUI MAIS C'EST SUR
    BEN OUI STU VEUX
    MAIS MAIS TU SAIS
    MAIS OUI HEIN
    EH OUAI MAIS DIS DONC
A schematic representation of this framework, including the basic terms, is given in (30).

<table>
<thead>
<tr>
<th>(30)</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
<th>(8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AH</td>
<td>BON</td>
<td>BEN</td>
<td>OUI/OUAIS</td>
<td>BEN</td>
<td>OUI/OUAIS</td>
<td>TU VOIS</td>
<td>HEIN</td>
<td></td>
</tr>
<tr>
<td>OH</td>
<td>MAIS</td>
<td>NON</td>
<td>MAIS</td>
<td>NON</td>
<td>OUAIS</td>
<td>TU SAIS</td>
<td>MAIS</td>
<td></td>
</tr>
<tr>
<td>EH</td>
<td>SI</td>
<td>ET</td>
<td>SI</td>
<td>ET</td>
<td>MAIS</td>
<td>S'TU VEUX</td>
<td>MAIS</td>
<td></td>
</tr>
<tr>
<td>HM</td>
<td>D'ACCORD</td>
<td>OU</td>
<td>D'ACCORD</td>
<td>BIEN SUR</td>
<td>BIEN SUR</td>
<td>ATTENDS</td>
<td>BIEN</td>
<td></td>
</tr>
<tr>
<td>etc.*</td>
<td></td>
<td>F'IN</td>
<td>F'IN</td>
<td>F'IN</td>
<td>F'IN</td>
<td>ECOUDE</td>
<td>F'IN</td>
<td></td>
</tr>
<tr>
<td></td>
<td>J'SAIS</td>
<td>DONC</td>
<td>J'SAIS</td>
<td>J'SAIS</td>
<td>J'SAIS</td>
<td>J'SAIS</td>
<td>J'SAIS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>C'EST SUR</td>
<td>C'EST SUR</td>
<td>C'EST SUR</td>
<td>C'EST SUR</td>
<td>C'EST SUR</td>
<td>C'EST SUR</td>
<td>C'EST SUR</td>
<td></td>
</tr>
<tr>
<td></td>
<td>+ positive adjective</td>
<td>+ positive adjective</td>
<td>+ positive adjective</td>
<td>+ positive adjective</td>
<td>+ positive adjective</td>
<td>+ positive adjective</td>
<td>+ positive adjective</td>
<td></td>
</tr>
<tr>
<td></td>
<td>J'SAIS PAS</td>
<td>J'SAIS PAS</td>
<td>J'SAIS PAS</td>
<td>J'SAIS PAS</td>
<td>J'SAIS PAS</td>
<td>J'SAIS PAS</td>
<td>J'SAIS PAS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>other</td>
<td>other</td>
<td>other</td>
<td>other</td>
<td>other</td>
<td>other</td>
<td>other</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(adj)verbs</td>
<td>(adj)verbs</td>
<td>(adj)verbs</td>
<td>(adj)verbs</td>
<td>(adj)verbs</td>
<td>(adj)verbs</td>
<td>(adj)verbs</td>
<td></td>
</tr>
<tr>
<td></td>
<td>of doubt'</td>
<td>of doubt'</td>
<td>of doubt'</td>
<td>of doubt'</td>
<td>of doubt'</td>
<td>of doubt'</td>
<td>of doubt'</td>
<td></td>
</tr>
</tbody>
</table>

* There are 23 terms in column (1), most of which occur only once.

* J'PENSE PAS / J'crois PAS / ÇA DÉPEND / PEUT-ÊTRE...

A number of methodological conventions were necessary in using this framework. Rapid repetitions of the same term(s) without prosodic interruption sometimes occurred. We coded these as single occurrences of the term, as in (31)–(33).
Treating each repetition separately would have required an otherwise unmotivated proliferation of columns in the framework.

Note that columns (3) and (5), and columns (4) and (6) are the pairs which contain similar lists of forms. If, however, a turn opening consisted of a single one of these forms or a pair of these, this was arbitrarily assigned to column (5) and/or (6) and not (3) and/or (4). Even if this single item was repeated as in (31) or (32) this was still considered as a single term as mentioned above. The choice was arbitrary but consistent. Finally, for statistical purposes, many similar forms were collapsed into one such as in (34).

\[
\begin{align*}
(34) \quad \text{d'accord} \\
\quad \text{j'suis d'accord} \\
\quad \text{tout-à-fait d'accord}
\end{align*}
\] = D'ACCORD 'I agree'

Before reporting on the statistical analysis, we can preview some of the results by sketching the contrastive functional roles of the different columns. That terms in the same column tend to have paradigmatically similar functions, many of which have been reported in earlier studies, is a gratifying validation of the distributional method – these functional similarities were not used as a criterion for constructing the columns.

4. The columns

Column (1) consists of what grammarians call interjections, such as \textit{ah}, \textit{oh} and \textit{eh}. Besides those listed we also found \textit{fioù}, \textit{berk}, \textit{bof}, \textit{tralala} and others. Together with \textit{bon} (column (2)) they constitute the immediate reaction to the preceding turn and are only minimally related to the current speaker's upcoming turn. Column (6) basically contains indications of agreement or disagreement with the propositional content of the preceding turn, while (4) serves the same role vis-à-vis the enunciative frame, or the speech act carried out by that turn. Columns (3) and (5) carry out the complex function of connecting the immediately preceding and following columns or turns. The final two columns, (7) and (8), contain a variety of tag-like expressions which are also found attached in an analogous way to sentence-final or sentence internal constituents, usually in a prosodically reduced form. These correspond in many ways to Vincent's (1983) notion of 'punctors'. Most of these are in column (7); column (8) only contains \textit{hein}. 

5. Statistical tools

Once the classification system was set up, all the 3700 turn openings in the corpus were coded. Note that in (30) with all the different possibilities in column (1) and in columns (4)–(7), there are literally millions of combinations theoretically possible. Of course most of these did not occur. Only about 200 different combinations did. To understand which forms tend to combine with each other and which do not, we are faced with a difficult task. Just looking at which combinations occur and which are absent is not very illuminating, especially since some of the individual forms are very rare and some very frequent, and this by itself may explain part of the occurrence pattern. Accordingly we devised two statistical procedures for dealing with these data.

The first is a method for detecting which combinations were really favoured or disfavoured. We denote by \( N(A) \) the number of form A which appear in column (1), by \( N(B) \) the number of form B which appear in column (2), etc. Note that A, B, etc. are very often the null term and then \( N(A) \) or \( N(B) \) is the number of times the appropriate slot was not filled. Then \( \frac{N(A)}{N} \) is the proportion of the total number \( N \) of openings which contain A, and is thus an estimate of the probability that A will appear in a given turn opening. To detect which combinations are favoured or disfavoured, we must demonstrate statistically that a certain 'null hypothesis' is untenable. This null hypothesis states that each slot is filled or empty independently of every other slot and that the form which is used to fill a slot does not affect what happens in neighbouring slots. Were this hypothesis of independence true, we could calculate the expected number of each combination of forms. For example, if a combination consists of form A plus form B plus ... plus form H, then:

\[
\text{(35) Expected number (form A \cdot form B \cdot \ldots \cdot form H)} = \frac{N(A)}{N} \times \frac{N(B)}{N} \times \ldots \times \frac{N(H)}{N}.
\]

By calculating these expected occurrences for all combinations and comparing them with the observed occurrences, we could detect which forms tended

\[^3\text{For combinations containing forms in columns (3) and/or (4), we use a somewhat more complicated calculation to ensure that the expected values respect our coding convention that column (4) can be filled only if column (5) or (6) is, and that column (3) can be filled only if column (4) or (5) is.}\]
to occur together more than expected and which tended to have co-occurrence restrictions.

A second statistical analysis was carried out to measure, for each form, to what extent it tends to stand alone or in one of a reduced number of combinations, versus to what extent it combines relatively freely with forms in other columns.

To calculate the expected number of different turn openings containing a given form X, we make use of the estimated probability of occurrence of all the forms in all the columns. Let \( p_A \) be the probability of occurrence of form A in its appropriate column, as estimated from \( N(A)/N \), and \( p_{0j} \) be the probability of no term in column j. Then for each possible turn opening T containing X, we can calculate the probability \( P(T) \) that any given turn will be opened by T. For example, if T is

**A --- X -- BC,**

then

\[
P(T) = p_A \times p_{02} \times p_{03} \times p_{04} \times p_X \times p_{06} \times p_B \times p_C
\]

and \( 1 - P(T) \) is the probability that any given turn will not be opened by T. Furthermore, \( (1 - P(T))^N \) is the probability that none out of N turns will be opened by T, so that \( 1 - (1 - P(T))^N \) is the probability that at least one of these N turns will be opened by T. It follows that the expected number of different turn openings containing X is a sum of form \( 1 - (1 - P(T))^N \) + \( 1 - (1 - P(U))^N \) + ... where T, U, ... are all the possible openings containing X. Then the number of different types of turn openings containing X actually observed, divided by the number expected, constitutes a measure of combinatoriality of the form X.

6. Quantitative results

A number of interesting facts emerge from the analysis of the data on how many slots are filled and how many are null, prior to examining the individual forms which occupy these slots:

- Two thirds of openings are non-null -- at least one slot is filled. This is quite remarkable in itself as it reflects an overwhelming difference between the spoken language and written varieties where turn openings are rare and even if they occur in some genres (theatre, screenplays) they usually tend to be constructed on a written model of dialogue;
— Combinations (4.5), (3.4.5), (3.4.5.6) occur more frequently than expected;
— There are many more turn openings made up of column (6) alone than expected, a consequence of which being that column (6) combines less with other columns than expected (e.g. column (4) and (6) co-occur but less frequently than expected);
— There is no co-occurrence of columns (3) and (5) alone;
— There is a clear pattern of column (1) and column (2) combining together;
— There are no other systematic co-occurrence tendencies or restrictions at the level of entire columns.

When we look at the results at the level of individual forms within two columns, the tendencies we have just described begin to be understandable. The tendency for column (6) to stand alone turns out to be due to the effect of *oui* (and its variant *ouais*). As we have already mentioned, in our data we included all turns including back channel activity and answers to yes/no questions. In these particular cases *oui* was the predominant form and it would usually occur by itself. When all such tokens were removed from the data *oui* loses its 'go it alone' character and behaves like all other forms when they occur alone, i.e. approximately as predicted.

The tendency for columns (1) and (2) to co-occur can be traced to a large number of *ah bon*. This is easily recognized by French speakers as a more or less frozen form, although it does not exclude cases of *ah* or *bon* co-occurring with other forms or alone. The form by form analysis revealed a number of other such expressions tending to become lexicalized (and considered as 'frozen expressions'). We compare the expected and observed numbers of these pairs in the following table:

<table>
<thead>
<tr>
<th>pairs</th>
<th>expected number</th>
<th>observed number</th>
</tr>
</thead>
<tbody>
<tr>
<td>eh ben</td>
<td>0.0</td>
<td>10</td>
</tr>
<tr>
<td>ah bon</td>
<td>2.1</td>
<td>26</td>
</tr>
<tr>
<td>bon ben</td>
<td>0.5</td>
<td>10</td>
</tr>
<tr>
<td>et pi</td>
<td>0.0</td>
<td>14</td>
</tr>
</tbody>
</table>

All of these are well known and it is a validation control of our method that each pair occurred far more frequently together than could be expected based only on the frequency of their component members.

These results are consistent with a tendency for these pairs to become lexicalized as a single entry, reduced prosodically. From the discourse analytic point of view the co-occurrence tendencies of columns (3), (4), (5) and (6) are far more interesting. The particular affinities and avoidance which occur in columns (3), (4), (5) and (6) form a rather complicated pattern. To simplify we will examine in table 1 a number of two-term combinations only.

Forms other than *mais* and *ben* occur in column (5) but we will only examine these two, being the most frequent ones. We include in 'mark of
agreement' such forms as *d'accord* and *j'suis d'accord* but also forms like *voilà bien sûr*... and all the *c'est*+positive adjective forms such as *c'est sur, c'est certain*, etc. There are, of course, nuances among these forms, but for this first overall analysis of combinatorial tendencies within turn openings, a finer categorization would have obscured more than it explained.

Note that there is a wide range of values in the third column of figures in table 1 representing the ratio of observed occurrences to what would have been expected based only on the overall productivity of the individual forms. Basically *mais* seems to prefer to follow the mark of agreement or disagreement (as in (c)) while *ben* does not (d). *Ben* rather seems to favour initial position (b) while *mais* does not and avoids introducing *oui* or other marks of agreement (a).

### Table 1.

<table>
<thead>
<tr>
<th></th>
<th>Observed</th>
<th>Expected</th>
<th>Observed/Expected</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Maïs oui</em></td>
<td>4</td>
<td>42</td>
<td>0.1</td>
</tr>
<tr>
<td><em>Maïs + mark of agreement</em></td>
<td>11</td>
<td>25</td>
<td>0.4</td>
</tr>
<tr>
<td><em>Maïs + mark of disagreement</em></td>
<td>18</td>
<td>13</td>
<td>1.4</td>
</tr>
<tr>
<td><em>Maïs non</em></td>
<td>22</td>
<td>28</td>
<td>0.8</td>
</tr>
<tr>
<td>(b)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Ben oui</em></td>
<td>24</td>
<td>13</td>
<td>1.8</td>
</tr>
<tr>
<td><em>Ben + mark of agreement</em></td>
<td>7</td>
<td>8</td>
<td>0.9</td>
</tr>
<tr>
<td><em>Ben + mark of disagreement</em></td>
<td>7</td>
<td>4</td>
<td>1.8</td>
</tr>
<tr>
<td><em>Ben non</em></td>
<td>10</td>
<td>9</td>
<td>1.1</td>
</tr>
<tr>
<td>(c)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Oui mais</em></td>
<td>114</td>
<td>49</td>
<td>2.3</td>
</tr>
<tr>
<td><em>Mark of agreement + mais</em></td>
<td>35</td>
<td>8</td>
<td>4.4</td>
</tr>
<tr>
<td><em>Mark of disagreement + mais</em></td>
<td>7</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td><em>Non mais</em></td>
<td>96</td>
<td>30</td>
<td>3.2</td>
</tr>
<tr>
<td>(d)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Oui ben</em></td>
<td>5</td>
<td>16</td>
<td>0.3</td>
</tr>
<tr>
<td><em>Mark of agreement + ben</em></td>
<td>0</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td><em>Mark of disagreement + ben</em></td>
<td>0</td>
<td>0.5</td>
<td>0</td>
</tr>
<tr>
<td><em>Non ben</em></td>
<td>0</td>
<td>10</td>
<td>0</td>
</tr>
</tbody>
</table>

How may we interpret this pattern? First recall that the roles of slots 3 and 4 are analogous to those of slots 5 and 6 respectively. *Oui* (and other marks of agreement) and *non* (and other marks of disagreement) may be a reaction to either the enunciative framework or to the informative content of the preceding turn. In most cases the positive reaction is toward the former (which explains why a speaker will often say 'yes' instead of 'no' when reacting to a preceding turn with negative informative content).

When slots 3, 4, 5 and 6 are all filled we postulate that the speaker reacts first to the enunciative framework, and then to the informative content, of the preceding turn. When only 5 and/or 6 are filled, there is some ambiguity as to whether the reaction is to the enunciative framework or the informative
content. This is due to our methodological convention of filling slots 3 and/or 4 only if forced to by the presence of material in slots 5 and/or 6. Thus when *oui* occurs alone in column (6), for example, it could be in reaction to informative content, but it is more often a positive reaction to the enunciative framework of the preceding turn.

Thus (c) and (d) unambiguously contain a reaction of agreement or disagreement with the preceding utterance followed by a connective form *mais* or *ben* introducing the speaker's own statement. (a) and (b) contain the same *mais* or *ben*, but sometimes introducing a reaction to the preceding turn's enunciative framework, and sometimes to its informative content.

(b) and (c) show two strategies for speakers to present their own attitudes or opinions, while mitigating the appearance of interactional conflict. In (c) the speaker first reacts to the enunciative framework involved in the preceding turn, thus removing the focus from the fact that in his/her own turn some sort of opposing conclusion (introduced by *mais*) will be presented. *Mai*s introduces a statement whose conclusion is opposed to the conclusion of the preceding turn or situation (Ducrot (1980)). By first dealing with the pragmatic appropriateness of the preceding turn before engaging in a turn which might be considered conflictual because of the opposite conclusions to be drawn in the sequence following *mais*, *oui mais* and *non mais* represent concessive strategies. They facilitate the sequencing of turns containing material which is potentially disruptive to the interaction because of incompatible informational claims or argumentative goals. Another concessive strategy involves *ben* as in (b). The reaction represented by *ben* does not connote as strongly as *mais* that what follows is in opposition to what precedes. *Ben*, in paradigmatic contrast to *mais*, is a more indirect way of introducing anything which could be construed as a disagreement or a mitigating agreement in setting up the next turn. *Ben* contains in its semantic description a distancing from the enunciative framework of the preceding speaker, implying that what he/she said is either irrelevant or too obvious. *Ben* is a prototypically concessive particle (see Anscombre (1985) for a general schema of concession) containing the hint of disagreement within an expression of basic agreement (e.g. *ben oui*, *ben d’accord*).

This brief description of the two forms (for more details see Ducrot (1980) and Vicher (forthcoming)) can explain why some strategies are favoured and others avoided. Conversation is based on a cooperation principle (Sacks, Schegloff and Jefferson (1974), Grice (1975)). If conversation is to continue, concession is favoured over direct opposition as a strategy for presenting one's ideas. Although we do find *mais* occurring alone (number observed equals number expected), *oui mais*, a concessive rather than a directly oppositional strategy, seems to be favoured. An analogous observation can be made about mark of agreement + *mais*. Agreement of some sort preceding a statement of opposition is a good concessive strategy for advancing one's own
viewpoint without being aggressive. *Oui mais* and mark of agreement + *mais* implement this strategy. *Non mais*, although superficially very different, actually contains the same argumentative movement. The *non* is usually a reaction to the enunciation of the preceding turn seen as a reply to what the present speaker had said in a penultimate turn, or to the conclusions the interlocutor drew from that turn.

From this description of *mais* we can understand why *mais oui* and *mais* + mark of agreement appear less often in conversation than expected. An initial *mais* is confrontational. Followed by *oui* (or other mark of agreement) it is, paradoxically, even more confrontational. Starting an utterance with *mais oui* or *mais d'accord*, the speaker is claiming that she or he knows or is already aware of what the interlocutor has said in the preceding turn, thus denying that the latter has brought anything new to the conversation. The disruptive potential of this is obvious.

This explains why there are fewer *mais oui* and *mais d'accord* than expected, and why those that do occur, tend to do so after questions. The denial of appropriateness introduced by *mais* (conveyed here by *oui* or *d'accord*) pertains not to the entire enunciative framework, but only to its interrogative aspect. Exception is taken to the preceding speaker’s questioning something in the present speaker’s penultimate turn, or simply to the use of a question rather than a simple affirmation.

*Ben*, while accepting following marks of agreement or disagreement, does not accept a strong preceding mark such as *d'accord* or *c'est sur*. It rarely happens that a speaker has a strong reaction to the enunciative framework, followed by an ambivalent reaction to its content. On the other hand, *ben* accepts a following mark of agreement or disagreement. Distance having been taken from the enunciative framework of the preceding turn, saying ‘*yes*’ or ‘*no*’ will not be seen as total agreement or disagreement.

7. Combinability

The second of our statistical analyses was designed to measure the combinability of each form, by comparing the number of different turn openings in which that form appeared with the number which could have been expected based on the overall frequency of the form and a model of independent filling (or not) of the various slots. Table 2 shows the results of this analysis.

We note first that no form occurred in as many different openings as expected under the random model. The forms which came the closest to unconstrained combinability were those like *tu vois, tu sais, si tu veux* and *hein*. This may be interpreted in terms of the relative functional independence of columns (7) and (8) from the preceding columns. Forms which appear the most constrained include *oui* and *non* in column (6), and *hmhm* in column (1).
These results are easily understood in terms of our counting responses to yes-no questions with the turn openings, as well as back channel activity, where single terms tend to occur in isolation.

Table 2.

<table>
<thead>
<tr>
<th></th>
<th>Observed number of different combinations (O)</th>
<th>Expected number of different combinations (E)</th>
<th>Index of combinability (O/E)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Himhm</td>
<td>7</td>
<td>26.5</td>
<td>0.26</td>
</tr>
<tr>
<td>Oh/Ah/Eh</td>
<td>41</td>
<td>68.2</td>
<td>0.60</td>
</tr>
<tr>
<td>Other</td>
<td>1</td>
<td>11.2</td>
<td>0.09</td>
</tr>
<tr>
<td>(2)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bon</td>
<td>17</td>
<td>34.9</td>
<td>0.49</td>
</tr>
<tr>
<td>(3)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mais</td>
<td>12</td>
<td>27.9</td>
<td>0.43</td>
</tr>
<tr>
<td>Ben</td>
<td>17</td>
<td>21.8</td>
<td>0.78</td>
</tr>
<tr>
<td>(4)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ou(a)j</td>
<td>38</td>
<td>69.9</td>
<td>0.54</td>
</tr>
<tr>
<td>Agreement</td>
<td>10</td>
<td>22.6</td>
<td>0.44</td>
</tr>
<tr>
<td>Non/si</td>
<td>25</td>
<td>51.4</td>
<td>0.49</td>
</tr>
<tr>
<td>Disagreement</td>
<td>3</td>
<td>6.8</td>
<td>0.44</td>
</tr>
<tr>
<td>(5)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ben</td>
<td>34</td>
<td>52.5</td>
<td>0.65</td>
</tr>
<tr>
<td>Mais</td>
<td>65</td>
<td>100.4</td>
<td>0.65</td>
</tr>
<tr>
<td>Et/ou</td>
<td>16</td>
<td>36.9</td>
<td>0.43</td>
</tr>
<tr>
<td>Donc</td>
<td>2</td>
<td>3.8</td>
<td>0.53</td>
</tr>
<tr>
<td>Fin</td>
<td>5</td>
<td>6</td>
<td>0.84</td>
</tr>
<tr>
<td>(6)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ou(r)j</td>
<td>13</td>
<td>57.7</td>
<td>0.23</td>
</tr>
<tr>
<td>Agreement</td>
<td>24</td>
<td>45.6</td>
<td>0.52</td>
</tr>
<tr>
<td>Non/si</td>
<td></td>
<td>48.5</td>
<td>0.23</td>
</tr>
<tr>
<td>Disagreement</td>
<td>24</td>
<td>31</td>
<td>0.77</td>
</tr>
<tr>
<td>Pi/Alors</td>
<td>22</td>
<td>27.9</td>
<td>0.78</td>
</tr>
<tr>
<td>Enfin/pourtant</td>
<td>5</td>
<td>10.1</td>
<td>0.49</td>
</tr>
<tr>
<td>(7)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tu vois</td>
<td>23</td>
<td>23.2</td>
<td>0.99</td>
</tr>
<tr>
<td>Tu sais</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Si tu veux</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tiens</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Donc</td>
<td>14</td>
<td>2.8</td>
<td>0.67</td>
</tr>
<tr>
<td>Ecoute</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attends</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other punctors</td>
<td>5</td>
<td>9.6</td>
<td>0.52</td>
</tr>
<tr>
<td>(8)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hein</td>
<td>21</td>
<td>23.7</td>
<td>0.89</td>
</tr>
</tbody>
</table>

We may next note that the terms in columns (3) and (4) tend to combine in a less diverse way than those in columns (5) and (6) (aside: from oui and non), probably due to the constraint that all turns involving the former two columns must, as a result of our coding conventions, also contain something
in one or both of the latter two columns. *Ben* in column (3) is an exception, as can be understood from the previous section where this form was found to occupy initial position of turns based on columns (3) to (6) much more easily than *mais*.

Finally, we note that *ah, oh* and *eh* show a slightly elevated index of combinability. This would have been even higher had it not been for an unusually large number of 'frozen' or almost frozen combinations like *ah bon, eh ben, ah oui* and *ah non*. Similarly, where it not for expressions like *bon ben* and *ah bon*, the combinability of *bon* would be higher than it is. Thus though our figures do not confirm a degree of independence of columns (1) and (2) with respect to what follows, in analogy to the independence of columns (7) and (8), we might expect a more refined calculation, taking into account the effect of frozen expressions, to reveal such a pattern.

The overall picture of the patterns of combinability of the columns is consistent with a 'constituent structure' of turn openings such as is displayed in (37).

(37)

8. Extralinguistic factors

Before concluding, we comment on the situational influences on turn openings. Despite our limited range of factors – debates versus conversations and native speakers versus fluent second language speakers – a number of interesting results emerge.

First, turn openings are generally more diverse in conversations than debates. Back channel activity was largely confined to *oui* in debates while the conversations also made liberal use of *hmhm* and *ah bon*. In debates no use was made of forms of agreement stronger than a plain *bon*, while these occurred several times in conversations.

Second, while non-native speakers adopt the same general strategy as native speakers to present an argument, i.e. the major tendencies are the same, some differences are to be noted. Native speakers have more elaborate turn openings than non-natives. The native speakers' combinations are longer. With non-natives, no combination exceeds more than three slots and only a few three-term combinations occur.
The native speakers' combinations are more diverse. Their data contain more than three times as many different types of opening than the non-natives. Proportionately speaking, the use of *oui* by itself plays a much more important role for the non-natives while the concessive strategy is largely confined to *oui mais*. One reason for this is the virtual absence of the form *ben* from the non-native data. This form, whose subtle functioning has remained mysterious and controversial to discourse analysts, has also proved refractory to second language learners.

9. Conclusion

In conclusion, we discuss how some of our results may be interpreted within the integrated pragmatic theory of enunciation developed by Anscombre and Ducrot (1983). The high rates of usage of *oui mais* and of *ben oui*/*ben non*, for example, can be understood in terms of the principle of cooperation underlying conversation: Turns opened by *oui mais* and *ben oui* are concessive and concession is a key rhetorical tool in conversation and debate.

But why these connective particles are used, and why one (*mais*) and not the other (*ben*) can be preceded by an agreement/disagreement particle, can only be explained, as we have tried to show, through a semantic description of *mais* and *ben*, taking into account the argumentative dynamics of language.

Finally an apology for the three-way methodological hybridization of distributionalism, variation theory and discourse analysis. It might seem that the discourse analytic approach, which starts from some surface form and attempts to distinguish its different functions and implications through examining its use in context, should be antithetical to the variationist practice of starting from a fixed underlying meaning, function or role, and trying to characterize the surface mechanisms which accomplish it. Indeed the latter may be seen as a denial of the type of distinction searched for in discourse analysis. In this study we tried to achieve a reconciliation through the intermediary of a theoretically neutral and, ideally, mechanical distributional analysis. Thus once the distributional analysis has characterized the slots or columns as well as their occupants we may precede with an essentially variationist analysis where slot membership is the underlying fixed entity. Once this is done the patterns of favoured versus disfavoured combinations represent a most appropriate type of phenomenon for a discourse analytic interpretation. The variants of a variable need not have a common referential meaning nor an identical pragmatic function and deep pragmatic distinctions are not inaccessible to statistical inference.
References


Vicher, A., forthcoming. Université de Paris V.