

Fréquences verbales dans les langues romanes

Autor(en): **Guitar, Henri**

Objektyp: **Article**

Zeitschrift: **Revue de linguistique romane**

Band (Jahr): **35 (1971)**

Heft 139-140

PDF erstellt am: **28.04.2024**

Persistenter Link: <https://doi.org/10.5169/seals-399509>

Nutzungsbedingungen

Die ETH-Bibliothek ist Anbieterin der digitalisierten Zeitschriften. Sie besitzt keine Urheberrechte an den Inhalten der Zeitschriften. Die Rechte liegen in der Regel bei den Herausgebern.

Die auf der Plattform e-periodica veröffentlichten Dokumente stehen für nicht-kommerzielle Zwecke in Lehre und Forschung sowie für die private Nutzung frei zur Verfügung. Einzelne Dateien oder Ausdrucke aus diesem Angebot können zusammen mit diesen Nutzungsbedingungen und den korrekten Herkunftsbezeichnungen weitergegeben werden.

Das Veröffentlichen von Bildern in Print- und Online-Publikationen ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. Die systematische Speicherung von Teilen des elektronischen Angebots auf anderen Servern bedarf ebenfalls des schriftlichen Einverständnisses der Rechteinhaber.

Haftungsausschluss

Alle Angaben erfolgen ohne Gewähr für Vollständigkeit oder Richtigkeit. Es wird keine Haftung übernommen für Schäden durch die Verwendung von Informationen aus diesem Online-Angebot oder durch das Fehlen von Informationen. Dies gilt auch für Inhalte Dritter, die über dieses Angebot zugänglich sind.

FRÉQUENCES VERBALES DANS LES LANGUES ROMANES

Nos études sur les corrélations de signifiants et de signifiés dans les langues romanes (*Travaux de Linguistique et Littérature*, 1969, VII, 131 ; *Miscellanea Barcinonensia*, 1970, XXV, 83) nous ont amené à dresser des listes de fréquence pour les grandes langues romanes : français, provençal, castillan, catalan, portugais, italien et tout récemment roumain (Mémoire de M^{lle} J. Martin).

Il nous a semblé intéressant d'utiliser ces listes pour préciser les fréquences d'emploi verbales, et rechercher comment elles varient d'une langue à l'autre. Nous ne retiendrons que les verbes dont la probabilité d'usage atteint $4 \cdot 10^{-5}$ ou, si l'on préfère, qui apparaissent au moins 4 fois dans un corpus de 100 000 mots. C'est cette seconde échelle que nous substituerons volontiers à la « probabilité » : toutes les valeurs sont multipliées uniformément par 100 000, et l'on évite l'emploi des nombres fractionnaires ou de leurs logarithmes négatifs.

Dans un premier temps, nous dresserons deux tableaux donnant : d'une part le pourcentage de verbes figurant dans le lexique de la langue, pourcentage détaillé pour les conjugaisons en *a*, en *e*, en *i* et pour le verbe *être* ; d'autre part, le pourcentage des formes de ces verbes rencontrées par rapport à l'ensemble des termes recensés dans les textes échantillonnés. Si nous nous référons à nos travaux précédents, le premier tableau correspondra au nombre N des vocables différents, le deuxième, à la masse M des occurrences recensées.

De la considération de ces deux tableaux, un certain nombre de remarques se dégagent immédiatement :

1^o Les verbes représentent sensiblement le quart du lexique d'une langue, et le cinquième des mots recensés. Cependant des fluctuations notables se manifestent autour de ces valeurs moyennes. Aux deux extrémités sont le français, avec 27,9 % des vocables et 22 % des occurrences, d'une part, et le roumain avec 19,1 % des vocables, ou l'espagnol avec 13,4 % des occurrences, d'autre part.

2° Le pourcentage des occurrences est toujours plus petit que celui des vocables. Cela signifie que les verbes sont moins employés que les autres mots du lexique ; leur fréquence moyenne est moindre.

Cependant la situation est renversée (sauf en espagnol) pour les verbes de la conjugaison en *e*, qui ont donc une fréquence d'emploi supérieure à la moyenne des vocables de la langue.

Au contraire, les verbes de la conjugaison en *a* ont une fréquence moyenne particulièrement faible, généralement inférieure à la moitié de la fréquence moyenne de l'ensemble des vocables : le pourcentage des occurrences n'atteint pas la moitié du pourcentage des vocables.

I. — *Tableau des vocables différents N %.*

	FRANÇAIS	PROVENÇAL	ESPAGNOL	CATALAN	PORTUGAIS	ITALIEN	ROUMAIN
Conjug. <i>a</i>	20,3	18,5	12,8	18,8	16,8	16,6	8,2
Conjug. <i>e</i>	4,6	2,9	4,6	2,9	3,9	6,4	4,4
Conjug. <i>i</i>	3,0	2,7	3,8	4,3	2,6	2,8	6,5
<i>Être</i>	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Total.....	27,9	24,1	21,2	26,0	23,3	25,8	19,1

II. — *Tableau des occurrences M %.*

	FRANÇAIS	PROVENÇAL	ESPAGNOL	CATALAN	PORTUGAIS	ITALIEN	ROUMAIN
Conjug. <i>a</i>	4,9	7,0	5,4	9,5	7,6	7,4	4,1
Conjug. <i>e</i>	10,2	5,9	3,2	6,3	5,2	8,3	8,2
Conjug. <i>i</i>	1,5	1,7	2,6	2,4	2,1	1,8	3,6
<i>Être</i>	5,4	3,4	2,2	1,8	2,1	2,7	2,7
Total.....	22,0	18,0	13,4	20,0	17,0	20,2	18,6

3° La conjugaison en *a* se caractérise par un nombre très important de vocables ; la conjugaison en *e* par une fréquence d'emploi souvent élevée. Ces deux éléments font simultanément défaut à la conjugaison en *i*, remarquable tant par le petit nombre des vocables que par celui des occurrences.

4° Le verbe *être* est très employé en français. Son usage est beaucoup plus réduit dans les langues ibériques ; la concurrence du verbe *estar* (compté avec les verbes en *a*) explique, dans une certaine mesure, cette divergence.

Le roumain, le provençal et l'italien occupent à cet égard une situation intermédiaire.

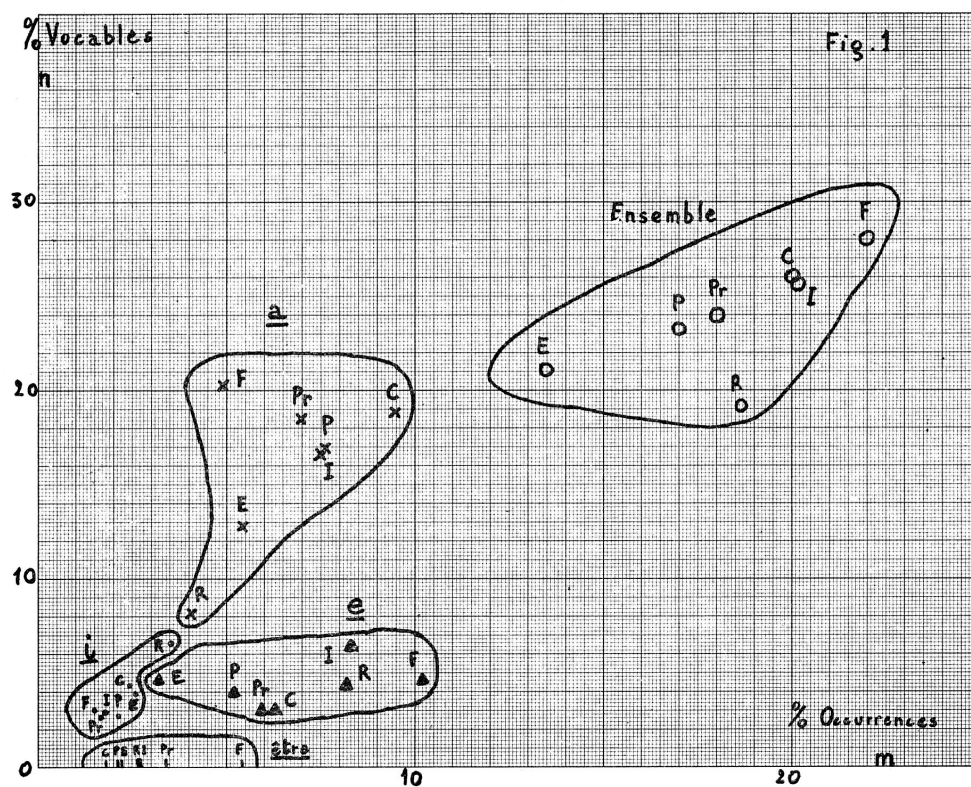
Sur la figure 1, nous avons porté en abscisse le pourcentage m des occurrences verbales, et en ordonnée celui n des vocables. Nous avons respectivement entouré par un trait : les points représentatifs de l'ensemble des verbes, les points représentatifs de la conjugaison en a , les points représentatifs de la conjugaison en e , ceux de la conjugaison en i , et ceux du verbe *être*. Chaque langue est représentée par sa seule initiale, sauf le provençal Pr , qu'il faut distinguer du portugais P .

Le français est très largement en tête pour l'abondance tant des vocables que des occurrences. Il est suivi par l'italien et le catalan, dont les points représentatifs sont presque confondus. Viennent ensuite le provençal et le portugais ; enfin l'espagnol, très en arrière, pratique un usage minimum du système verbal. Le roumain a moins de vocables verbaux que l'espagnol, mais il les emploie autant que le portugais et le provençal.

Les points représentatifs de la conjugaison en a se groupent dans une zone où n a des valeurs élevées et m des valeurs moyennes. L'espagnol et le roumain se signalent par un nombre minimum des vocables, le français par un nombre maximum (dû à l'abondance de néologismes comme *émotionner* pour *émouvoir*, *réceptionner* pour *recevoir*, *solutionner* pour *résoudre*, etc.) ; le catalan, par un nombre maximum d'occurrences. A l'intérieur du triangle défini par ces points, l'italien, le portugais et le provençal se rassemblent près du centre de gravité.

Les points représentatifs de la conjugaison en e s'étirent dans le sens des abscisses, pour des valeurs médiocres des ordonnées. Le français l'emporte de loin pour les occurrences, suivi par l'italien et le roumain. Le catalan, le provençal et le portugais ont des points très voisins, tandis que l'espagnol pratique encore l'emploi minimum.

Les points représentatifs de la conjugaison en i se serrent sur une aire exiguë, avec peu de vocables et peu d'occurrences. C'est vraiment, sauf en roumain, la parente pauvre de la famille verbale. Pour une fois, l'espagnol a la seconde place des occurrences (et la troisième des vocables), car il a favorisé la conjugaison en i : *decir* (ptg. *dizer*), *recibir* (ptg. *receber*), *hervir* (ptg. *ferver*), etc... Mais cette faveur est très relative, et, en espagnol aussi, la conjugaison en i occupe à tous points de vue le dernier rang. Nous ne pouvons donc souscrire à l'affirmation de Menéndez Pidal (*Manual de gramática histórica española* § III) : « La conjugación -IRE es la segunda en riqueza después de la -are ».



Enfin les points représentatifs du verbe *être*, qui ne correspondent qu'à un seul vocable, sont pratiquement sur l'axe des abscisses.

Il est, bien entendu, possible de faire intervenir, dans la présentation des résultats, les probabilités des diverses catégories verbales, ou, mieux, les fréquences moyennes sur un recensement de 100 000 mots (probabilité $\times 10^5$).

III. — *Tableau des fréquences moyennes.*

	FRANÇAIS	PROVENÇAL	ESPAGNOL	CATALAN	PORTUGAIS	ITALIEN	ROUMAIN
Conjug. <i>a</i>	11	12	16	18	12	12	21
Conjug. <i>e</i>	107	64	27	77	36	35	77
Conjug. <i>i</i>	25	19	27	19	20	17	23
<i>Être</i>	4 417	3 189	2 054	1 688	1 939	2 536	2 592
Ensemble.....	38	24	24	27	19	21	40

Sur la figure 2, nous avons représenté les pourcentages de vocables en fonction des fréquences moyennes correspondantes.

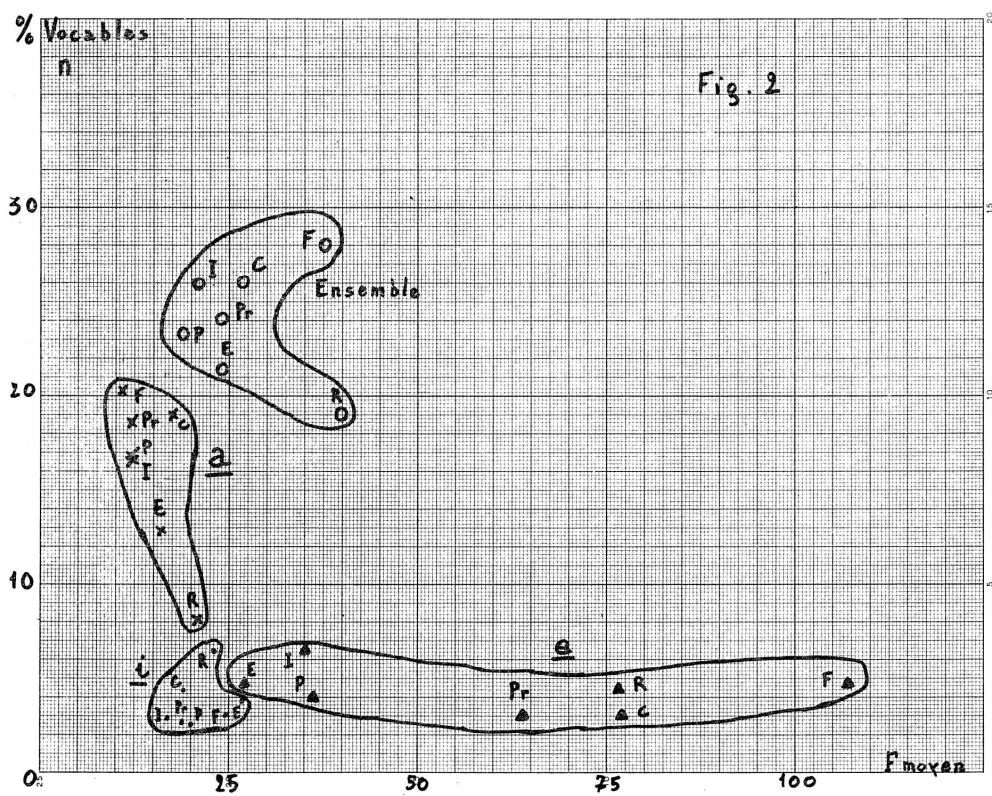
Les situations relatives des diverses aires rappellent celles de la figure 1. Cependant l'étirement des points représentatifs de la conjugaison en *e*, dans le sens de l'axe des abscisses, se trouve encore accusé, tandis que les points représentatifs de la conjugaison en *a*, et ceux de l'ensemble des verbes, se retirent vers les valeurs inférieures des abscisses. Les points représentatifs du verbe *être* auraient des abscisses qui atteindraient 8 mètres, et, par conséquent sont hors des limites de la figure.

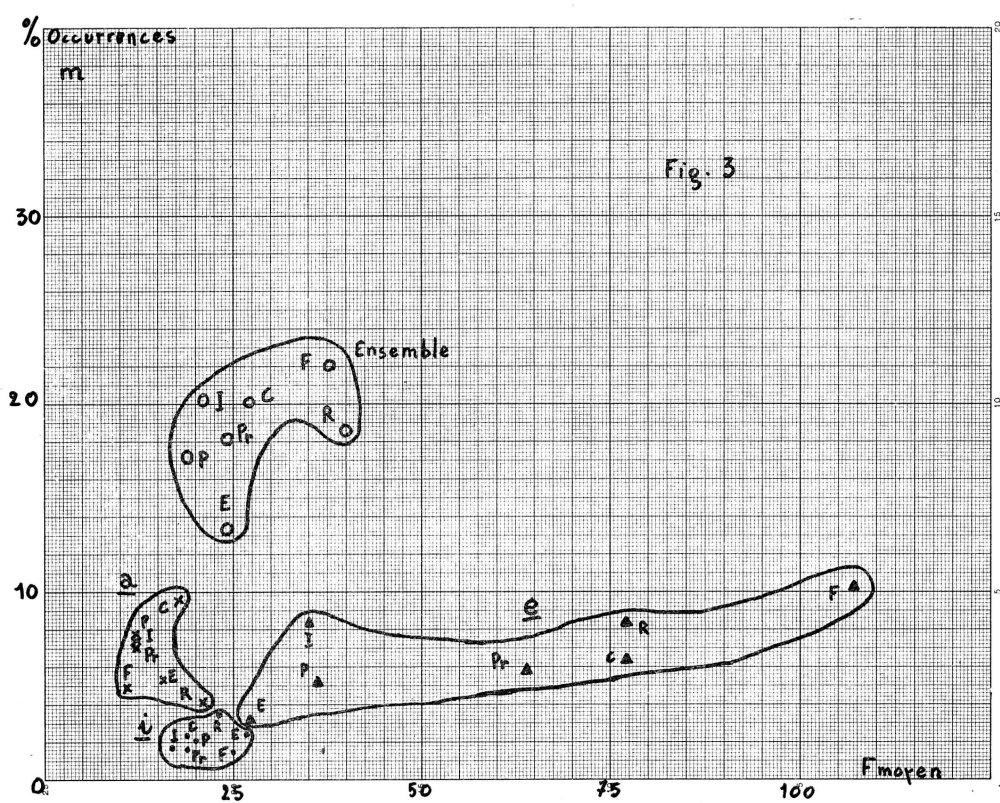
La figure 3 représente les pourcentages d'occurrences en fonction des valeurs moyennes de la fréquences pour chaque catégorie verbale. La répartition des aires appelle les mêmes remarques que dans le cas de la figure 2.

L'ensemble des trois graphiques constitue les projections sur les trois plans d'un trièdre trirectangle $[n, \bar{F}, m]$, des points représentatifs, dans l'espace, des diverses conjugaisons et de leur ensemble.

Leur rapprochement nous confirme l'exubérance du système verbal et de ses emplois en français, et, au contraire, sa modestie en espagnol et en roumain ; l'italien et le catalan, puis le provençal et le portugais occupent des situations intermédiaires.

La conjugaison en *a* est très riche en vocables différents, mais de fréquence relativement faible, surtout en français.





La conjugaison en *e* compte peu de vocables, mais leur fréquence moyenne d'emploi est très élevée, surtout en français.

La conjugaison en *i* ne compte pas plus de vocables que la conjugaison en *e*, mais leur fréquence d'emploi ne dépasse guère celle de la conjugaison en *a*.

C'est en espagnol et en roumain que les points représentatifs des trois conjugaisons sont les plus rapprochés, c'est à dire qu'elles sont relativement le plus équilibrées ; c'est en français qu'ils sont le plus dispersés.

*
* *

Après ces observations portant sur l'ensemble des vocables, nous nous proposons maintenant de rechercher la répartition des diverses conjugaisons selon les ordres de grandeur de la fréquence.

Nous allons répartir les vocables en quatre catégories dont nous nous fixons ainsi les limites :

1) ceux de très grande fréquence, dont la probabilité est supérieure ou égale à $50 \cdot 10^{-5}$ (fréquence supérieure ou égale à 50 dans un recensement de 100 000 mots).

2) Ceux de grande fréquence, dont la probabilité est supérieure ou égale à $20 \cdot 10^{-5}$, mais inférieure à $50 \cdot 10^{-5}$ (fréquence comprise entre 20 et 50 dans un recensement de 100 000 mots).

3) ceux de moyenne fréquence, dont la probabilité est supérieure ou égale à $10 \cdot 10^{-5}$, mais inférieure à $20 \cdot 10^{-5}$ (fréquence comprise entre 10 et 20 dans un recensement de 100 000 mots).

4) ceux de petite fréquence, dont la probabilité est supérieure ou égale à $4 \cdot 10^{-5}$, mais inférieure à $10 \cdot 10^{-5}$ (fréquence comprise entre 4 et 10 dans un recensement de 100 000 mots).

Nous laissons de côté les mots de très petite fréquence, dont la probabilité est inférieure à $4 \cdot 10^{-5}$.

Pour chacune de ces catégories, nous allons donner les pourcentages verbaux tant par rapport au nombre des vocables différents qu'au nombre des occurrences.

Les résultats de cette statistique sont consignés dans le tableau IV pour les vocables, et dans le tableau V pour les occurrences ; le tableau VI nous donne la fréquence moyenne correspondant à chacune des catégories précédentes.

Avec l'aide de ces trois tableaux, il nous est encore possible de dresser des graphiques. Nous nous occuperons tout d'abord de ceux qui correspondent

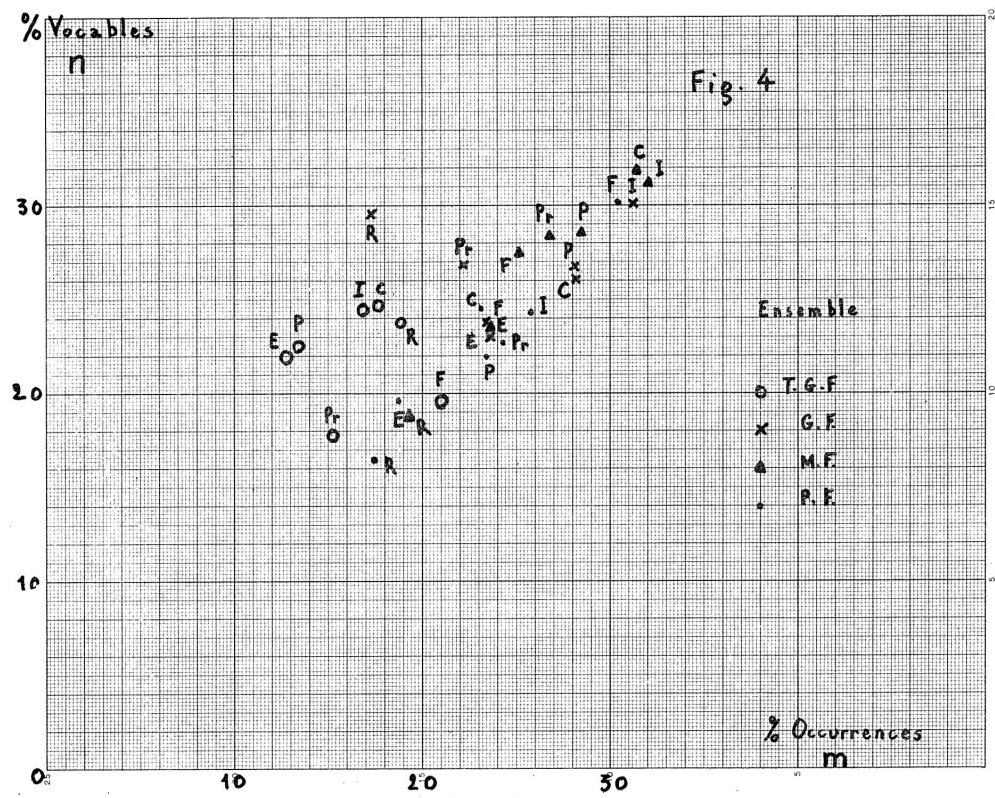
IV. *Tableau des vocables différents N %.*

		FRANÇAIS	PROVENÇAL	ESPAGNOL	CATALAN	PORTUGAIS	ITALIEN	ROUMAIN
Conjug. <i>a</i>	T. G. F.....	7,2	7,7	8,8	11,8	12,8	9,4	7,4
	G. F.....	11,9	18,8	13,2	17,3	13,5	15,8	13,3
	M. F.....	18,1	20,4	13,7	23,2	19,9	16,2	7,2
	P. F.....	24,4	18,9	12,8	18,6	17	17,4	7,9
Conjug. <i>e</i>	T. G. F.....	9,4	8,2	7,7	8,9	6,1	11,2	10,3
	G. F.....	7,6	4,9	4,8	4,7	8,2	11,0	9,7
	M. F.....	5,9	3,6	5,6	3,4	5,4	10,5	4,0
	P. F.....	3,2	1,7	3,7	1,6	2,9	4,7	2,4
Conjug. <i>i</i>	T. G. F.....	2,2	1,4	4,9	3,5	3,1	3,3	5,7
	G. F.....	4,3	3,2	5,1	4,1	4,4	3,3	6,6
	M. F.....	3,6	4,4	4,1	4,7	3,2	4,5	7,7
	P. F.....	2,7	2,2	3,2	4,3	2,1	2,4	6,1
<i>Être</i>	T. G. F.....	0,7	0,5	0,5	0,5	0,5	0,5	0,4
Total	T. G. F.....	19,5	17,8	21,9	24,7	22,5	24,4	23,8
	G. F.....	23,8	26,9	23,1	26,1	26,6	30,1	29,6
	M. F.....	27,6	28,4	23,4	31,9	28,5	31,2	18,9
	P. F.....	30,3	22,8	19,7	24,5	22,0	24,5	16,4

à l'ensemble des conjugaisons : la figure 4 situe les pourcentages de vocables en fonction des pourcentages d'occurrences ; la figure 5, les pourcentages de vocables par rapport aux fréquences moyennes ; la figure 6, les pourcentages d'occurrences par rapport aux fréquences moyennes.

Dans le cas présent, les fréquences moyennes varient de quelques unités à quelques centaines. Si nous les portons en valeurs naturelles, les faibles valeurs des abscisses ne se distingueraient pas entre elles ; nous sommes donc amené à adopter une échelle logarithmique.

L'observation de la figure 4 nous révèle que beaucoup de points représentatifs sont voisins de la bissectrice des axes de coordonnées. Ceci veut dire que les verbes présentent souvent un pourcentage de vocables sensiblement égal à leur pourcentage d'occurrences, donc que leur fréquence moyenne d'emploi est équivalente à celle des autres mots situés dans la même bande de fréquence. Mais ceci est inexact pour les verbes de très grande fréquence en espagnol, en catalan, en portugais, en italien et en roumain, ainsi que



V. Tableau des occurrences M %.

		FRANÇAIS	PROVENÇAL	ESPAGNOL	CATALAN	PORTUGAIS	ITALIEN	ROUMAIN
Conjug. <i>a</i>	T. G. F.....	1,8	2,5	2,8	6,3	3,9	3,5	2,6
	G. F.....	10,8	15,1	12,9	18,5	13,8	16,8	8,4
	M. F.....	16,7	19,2	13,4	23,2	18,8	16,6	7,6
	P. F.....	24,4	20,0	11,6	17,1	17,8	18,1	8,2
Conjug. <i>e</i>	T. G. F.....	11,3	7,1	5,2	7,2	5,1	8,4	9,5
	G. F.....	8,2	4,0	5,4	5,3	9,4	10,5	5,9
	M. F.....	5,5	3,5	4,7	3,7	5,4	11,1	3,9
	P. F.....	3,0	1,9	3,7	1,6	3,0	4,9	2,7
Conjug. <i>i</i>	T. G. F.....	1,1	1,0	1,9	1,7	1,5	1,1	2,8
	G. F.....	4,4	3,2	5,2	4,3	4,9	3,8	4,1
	M. F.....	2,9	4,1	5,4	4,5	3,3	4,4	7,8
	P. F.....	2,9	2,4	3,3	4,4	2,6	2,7	6,5
<i>Être</i>	T. G. F.....	6,8	4,8	2,9	2,4	2,9	3,8	3,7
Total	T. G. F.....	21,0	15,4	12,8	17,6	13,4	16,8	18,6
	G. F.....	23,4	22,1	23,5	28,1	28,1	31,1	17,4
	M. F.....	25,1	26,8	23,5	31,4	28,5	32,1	19,3
	P. F.....	30,3	24,1	18,6	23,1	23,4	25,7	17,4

pour les verbes de grande fréquence en provençal et en roumain ; leur pourcentage de vocables est supérieur à leur pourcentage d'occurrences : leur fréquence moyenne est donc inférieure à la fréquence moyenne des autres mots de la même bande de fréquence. C'est donc le seul poids de ces verbes de très grande fréquence qui nous avait amené à formuler notre deuxième observation sur l'ensemble du système verbal, à savoir que la fréquence moyenne des verbes est inférieure à celle des autres mots de la langue.

Dans l'emploi de ces verbes de très grande fréquence, nous constatons le voisinage des points représentatifs de l'espagnol et du portugais, d'une part, du catalan et de l'italien, d'autre part ; le français et le provençal sont l'un et l'autre isolés.

Dans le tableau des fréquences moyennes de l'ensemble des verbes, notre attention est attirée par la valeur relative très élevée de la fréquence des verbes de la bande de T. G. F. en français. Cette valeur est à peu le double

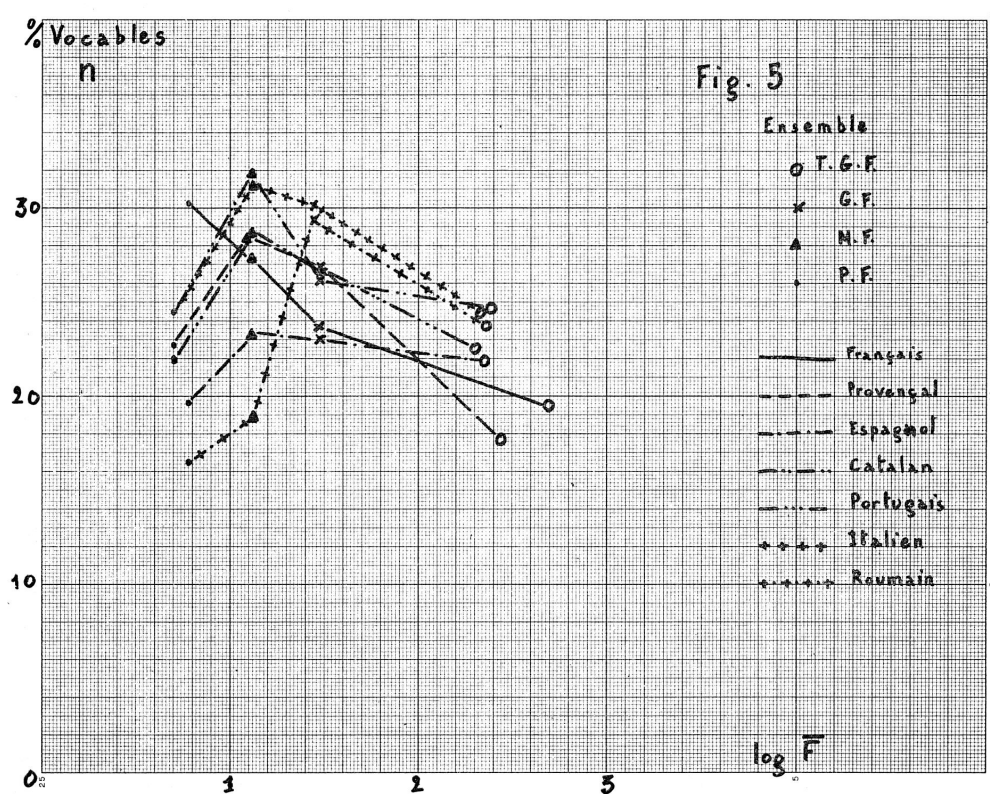
VI. — *Tableau des fréquences moyennes $\bar{F}$.*

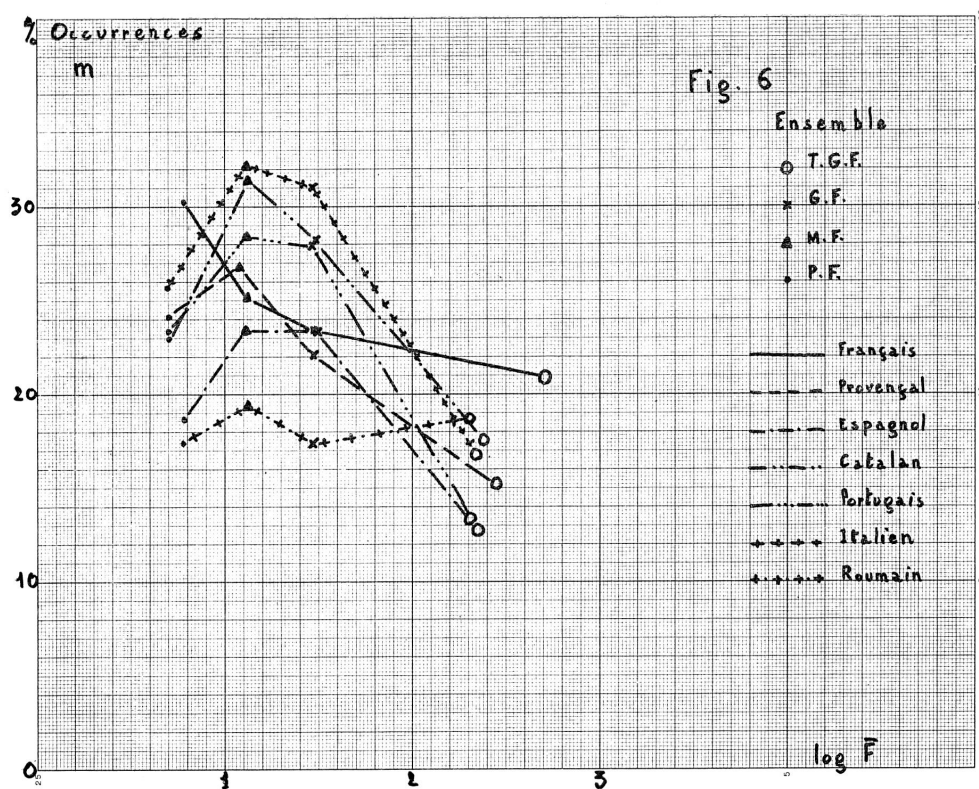
	FRANÇAIS	PROVENÇAL	ESPAGNOL	CATALAN	PORTUGAIS	ITALIEN	ROUMAIN
Conjug. <i>a</i> {							
T. G. F. . .	118	106	124	183	106	117	102
G. F.	26	28	28	29	29	29	28
M. F.	13	12	12	13	13	13	14
P. F.	6	5	5	5	5	5	6
Conjug. <i>e</i> {							
T. G. F. . .	565	281	261	280	285	232	267
G. F.	31	28	32	30	31	26	29
M. F.	13	12	11	14	13	14	13
P. F.	5	5	6	5	5	5	7
Conjug. <i>i</i> {							
T. G. F. . .	234	226	150	173	169	104	140
G. F.	30	34	29	29	31	31	24
M. F.	11	12	17	13	14	13	13
P. F.	6	5	6	5	5	5	6
<i>Être</i>	4 417	3 189	2 054	1 688	1 939	2 536	2 592
Ensemble {							
T. G. F. . .	505	279	226	247	205	215	225
G. F.	29	29	29	29	30	28	28
M. F.	13	12	13	13	13	13	13
P. F.	6	5	6	5	5	5	6

de ce qu'elle est dans les autres langues romanes. Le français fait donc un usage intense de ses verbes usuels.

L'examen des courbes représentatives des pourcentages de N ou de M en fonction de la fréquence, met encore le français à part. Les deux pourcentages sont l'un et l'autre, fonctions décroissantes des fréquences moyennes, alors que, pour toutes les autres langues nous obtenons des courbes à maxima, le maximum correspondant à la moyenne fréquence (et à la grande fréquence pour les vocables du roumain). C'est la masse de ces verbes de petite fréquence qui permet au français d'être « très largement en tête pour l'abondance tant des vocables que des occurrences », ainsi que nous le remarquons dans la première partie.

Les courbes représentatives du catalan et de l'italien sont les plus élevées ; celles de l'espagnol et du roumain sont les plus basses ; le portugais occupe une situation intermédiaire ; quant au provençal, il se rapproche du portugais pour les faibles fréquences, mais de l'espagnol, pour les grandes.





*
* *

Avec les données des tableaux IV, V et VI, relatives à la conjugaison en *a*, il nous est possible de construire les figures 7, 8 et 9.

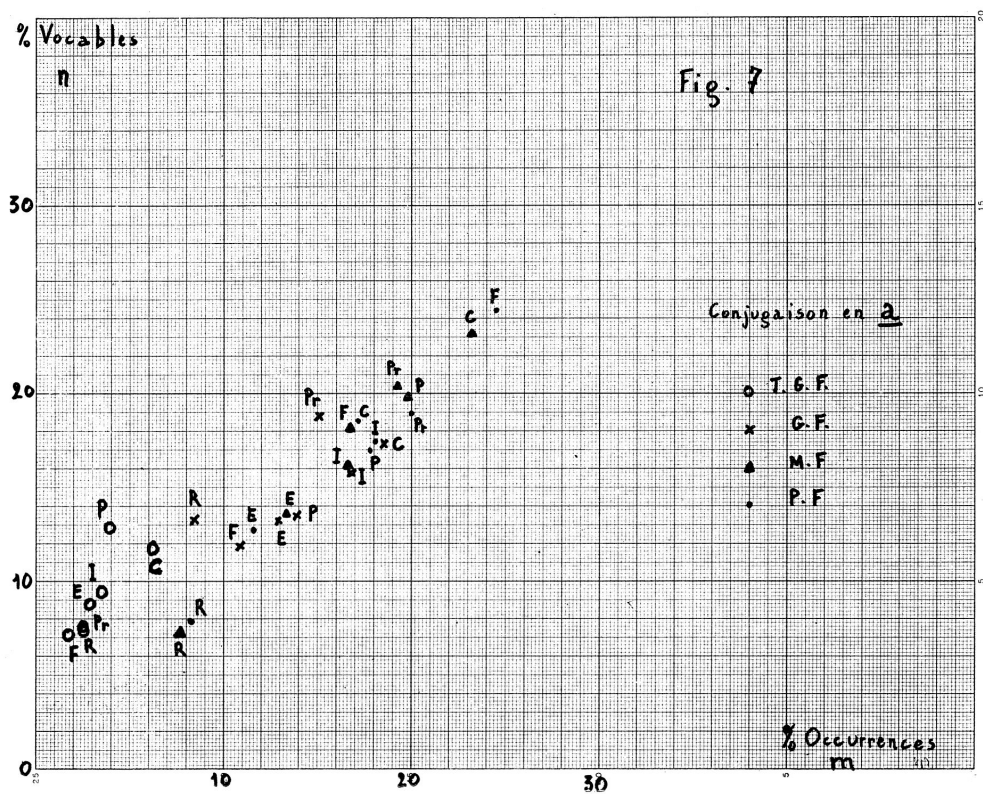
La figure 7 rappelle la figure 4, mais avec une diminution générale de 10, tant des ordonnées que des abscisses. Ici encore, beaucoup de points représentatifs sont voisins de la bissectrice des axes de coordonnées ; mais ils se rassemblent surtout entre les valeurs 10 et 20, et non plus entre les valeurs 20 et 30. Ici encore, les verbes de très grande fréquence ont un pourcentage de vocables très supérieur au pourcentage d'occurrences, par conséquent, une fréquence moyenne très inférieure à celle des autres mots de très grande fréquence ; ceci étant particulièrement marqué pour le portugais.

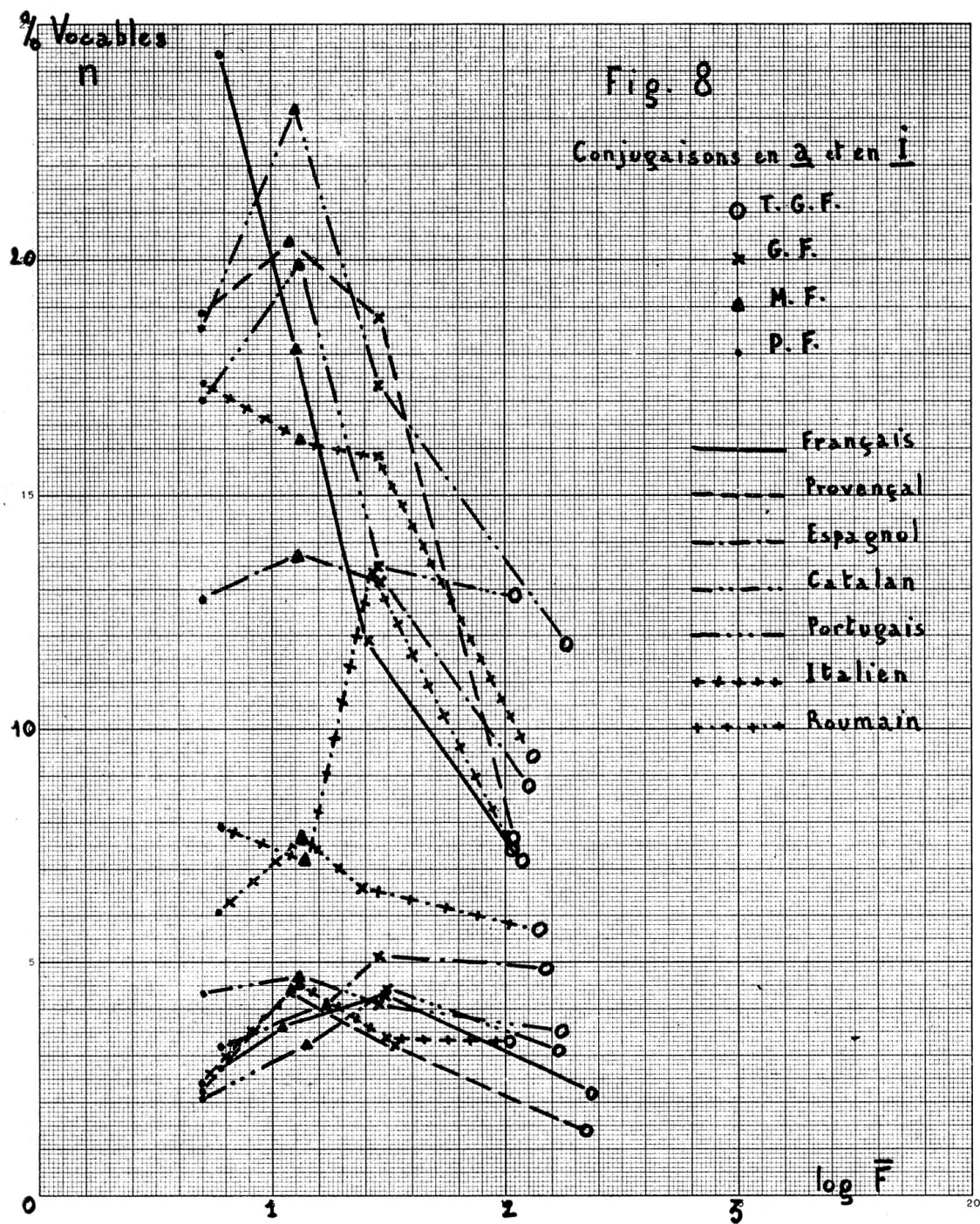
La conjugaison en *a* offre donc une répartition analogue à celle de l'ensemble des conjugaisons, dont elle représente en gros l'élément majoritaire, si ce n'est qu'il faut noter un effondrement particulier de la fréquence moyenne dans la bande des très grandes fréquences.

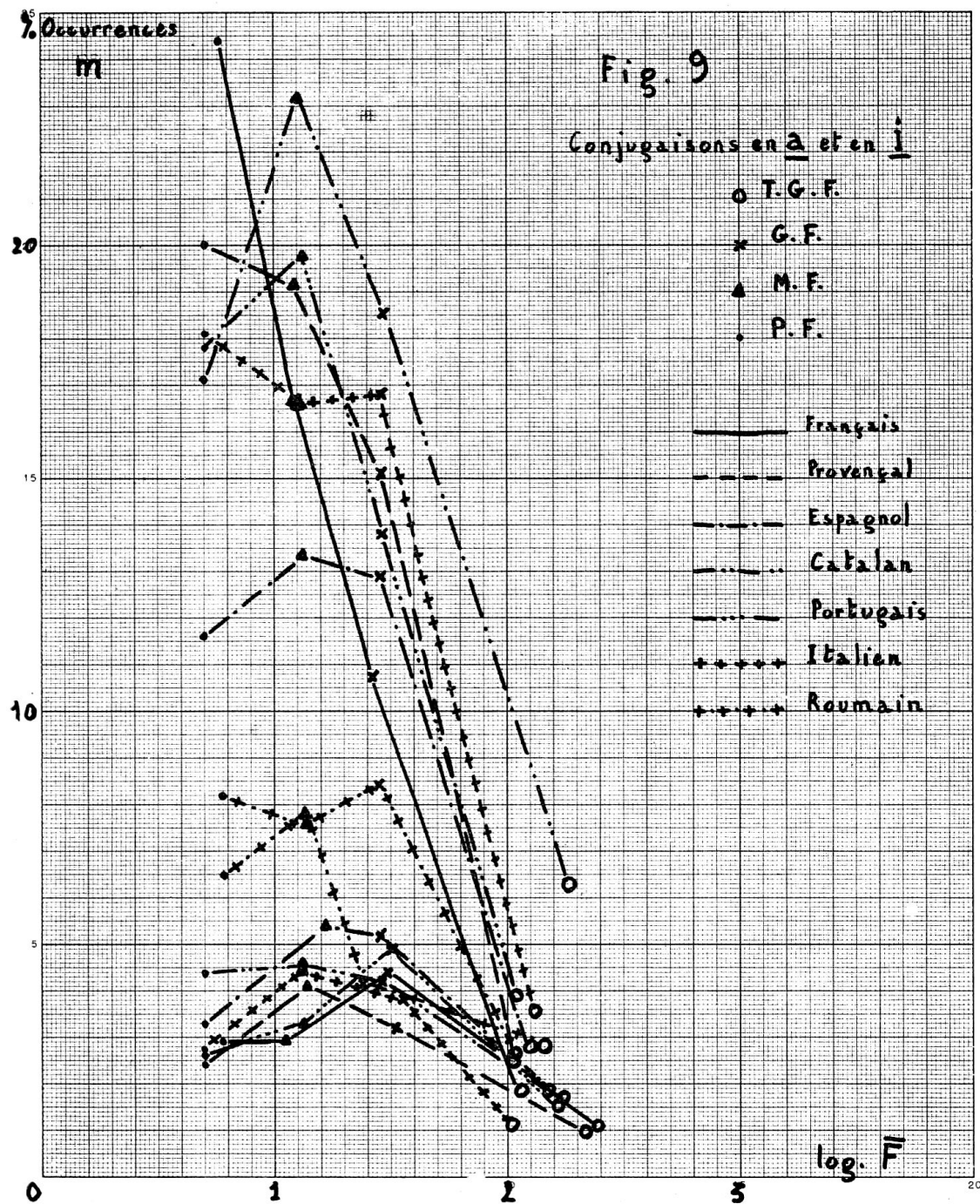
Les figures 8 et 9 appellent pour le français des remarques analogues à celles faites sur l'ensemble des conjugaisons : aussi bien les pourcentages de vocables que ceux d'occurrences y sont fonction décroissante de la fréquence ; dans les deux cas, les courbes du français partent du point le plus élevé parmi les petites fréquences pour aboutir au point le plus bas parmi les très grandes fréquences. Dans la bande de petite fréquence du français, près d'un quart de l'ensemble des vocables et de leurs occurrences correspondent à des verbes de la conjugaison en *a* ; au contraire, dans la bande de très grande fréquence, les verbes en *a* ne représentent qu'un vocable sur quatorze et leurs emplois ne correspondent même pas à une occurrence sur cinquante.

Mais, dans la conjugaison en *a*, l'isolement du français n'est plus aussi absolu que dans l'ensemble des conjugaisons. L'italien présente lui aussi des fonctions décroissantes de la fréquence, quoique avec des variations beaucoup moins accusées que le français. Dans la bande de petite fréquence, les verbes de la conjugaison en *a* se manifestent sensiblement une fois sur sept, donc, près de deux fois moins qu'en français ; dans la bande de très grande fréquence, ils arrivent à représenter un vocable sur dix, et une occurrence sur trente.

Le provençal prend aussi une allure comparable à celle du français, mais uniquement pour les occurrences. De une sur cinq pour les petites fréquences,







les occurrences de la conjugaison en *a* tombent à une sur quarante pour les très grandes fréquences.

En ce qui concerne les pourcentages de vocables, le provençal présente une courbe de même type que les langues ibériques avec un maximum accusé sur la moyenne fréquence.

Les trois langues ibéroromanes conservent pour la conjugaison en *a* des courbes analogues à celles obtenues pour l'ensemble des conjugaisons. Tant pour les vocables que pour les occurrences, nous obtenons un pourcentage maximum dans la bande des moyennes fréquences.

L'espagnol présente un maximum peu accusé, et sa courbe ne monte jamais bien haut : ceci ne nous étonne pas puisque la première partie de cette étude nous avait montré la pauvreté relative de sa conjugaison en *a*.

Le catalan et le portugais se rapprochent beaucoup, avec l'un et l'autre des maxima très accusés, le catalan se signalant néanmoins par sa plus grande richesse verbale.

Le roumain se cantonne dans des pourcentages très bas, avec maximum pour la grande fréquence.

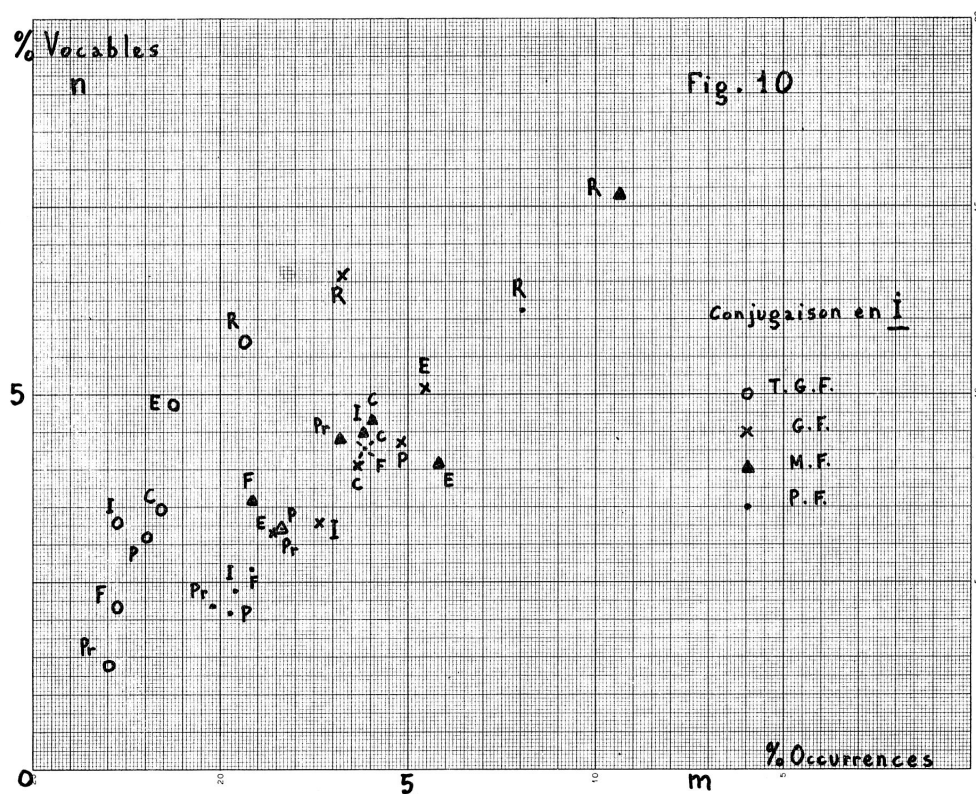
*
* *

Sur les figures 8 et 9, nous avons porté en outre les courbes représentatives de la conjugaison en *i*. Ceci nous a été possible grâce à la faiblesse des ordonnées par rapport à celles correspondant à la conjugaison en *a* ; nous n'avons été légèrement gêné que pour la bande de très grande fréquence du graphique des occurrences.

En revanche, nous avons dû représenter à part, sur la figure 10, et avec une échelle quatre fois plus grande que précédemment, les points donnant les pourcentages de vocables en fonction des pourcentages d'occurrences ; encore sont-ils souvent très proches les uns des autres.

Un examen sommaire de cette figure 10 nous montre toujours la majorité des points représentatifs groupés près de la bissectrice, avec une disposition rappelant celle de la figure 7 ; seulement l'échelle est quatre fois plus grande : le volume de la conjugaison en *i* est donc à peine le quart de celui de la conjugaison en *a*.

Cette observation cesse aussi d'être exacte en ce qui concerne la bande de très grande fréquence ; mais ici les points représentatifs s'écartent beaucoup plus de la bissectrice que dans le cas de la conjugaison en *a*. Le pourcentage des vocables est relativement très élevé en italien et dans les langues ibériques, surtout en espagnol.



Nous avons déjà noté cette faveur accordée par l'espagnol à la conjugaison en *i*, même pour ses verbes usuels, c'est à dire de très grande fréquence : dans cette bande, le pourcentage des occurrences espagnoles de la conjugaison en *i* est supérieur à celui des occurrences françaises de la conjugaison en *a*. Mais c'est en roumain que la conjugaison en *i* connaît son plus grand développement, surtout quant au nombre des vocables.

L'examen des courbes représentatives de la conjugaison en *i* sur les figures 8 et 9 appelle quelques observations. Ici toutes les courbes présentent un maximum, mais il se situe tantôt sur la moyenne fréquence (comme dans les cas précédents), tantôt sur la grande fréquence. Le provençal, le roumain, le catalan et l'italien ont leurs maxima sur la moyenne fréquence ; le français et le portugais, sur la grande fréquence ; l'espagnol sur la moyenne fréquence pour les occurrences, et sur la grande fréquence pour les vocables.

*
* *

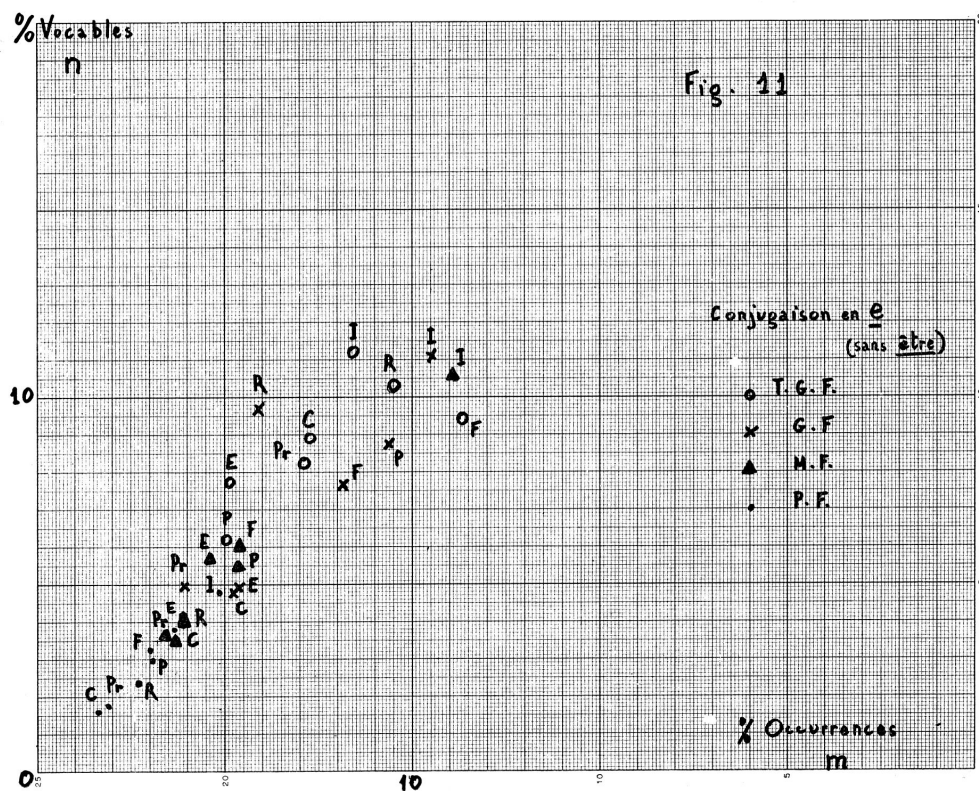
Si la conjugaison en *i* a pu nous donner l'impression d'une « mini-conjugaison en *a* », il n'en est pas de même avec la conjugaison en *e*. Les répartitions des verbes par bandes de fréquences sont structurées de façon totalement différente.

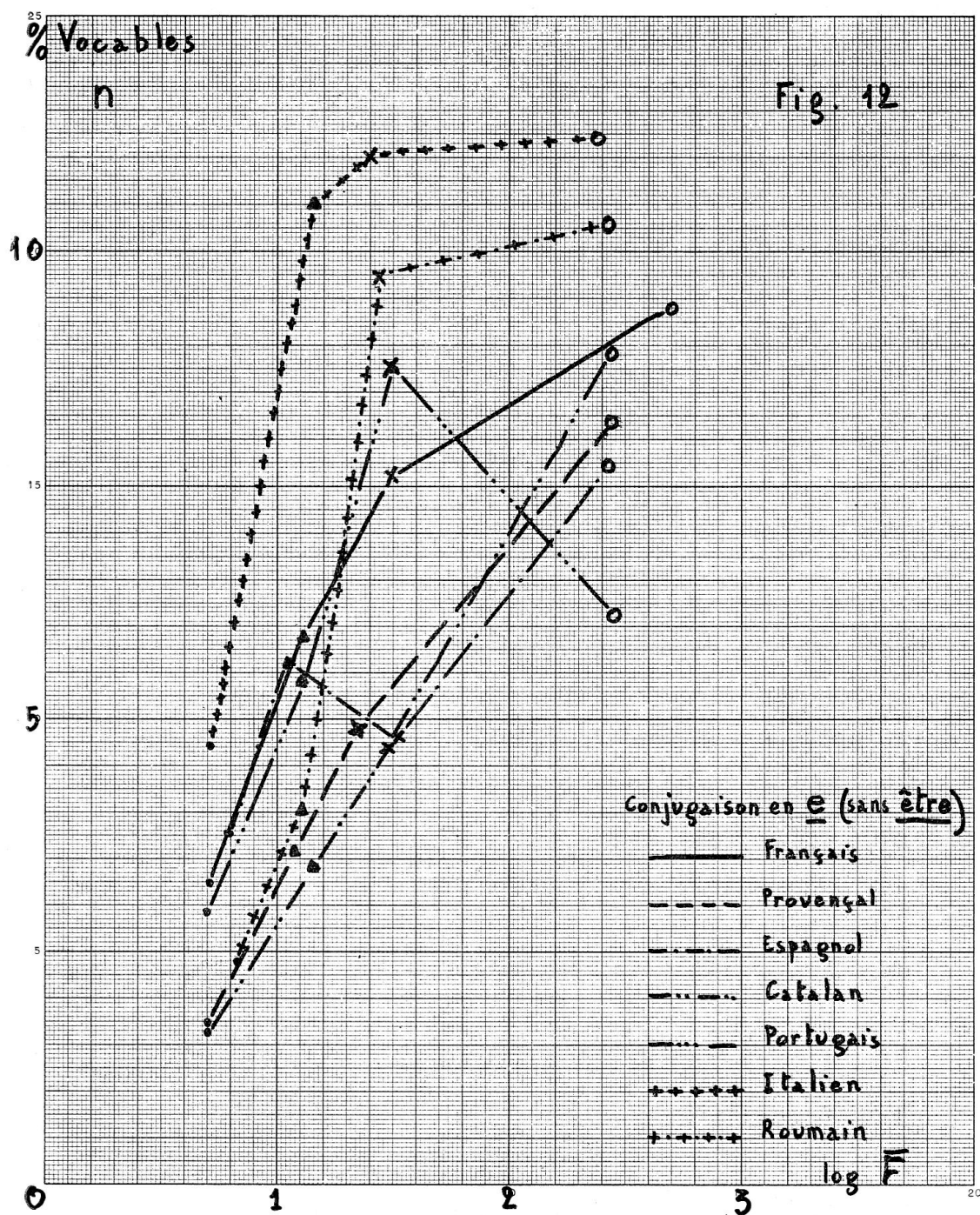
Fallait-il introduire le verbe *être* dans les décomptes relatifs à la conjugaison en *e* ? Morphologiquement cette inclusion s'imposerait ; mais le rôle syntactique du verbe *être* est tellement particulier que nous avons hésité. En définitive nous avons décidé de présenter les deux séries de graphiques, ceux des figures 11, 12, 13 avec le verbe *être* exclu, ceux des figures 14, 15, 16 avec le verbe *être* inclus. Bien entendu, l'inclusion du verbe *être* entraîne une augmentation de la fréquence moyenne dans la bande de très grande fréquence ; elle devient :

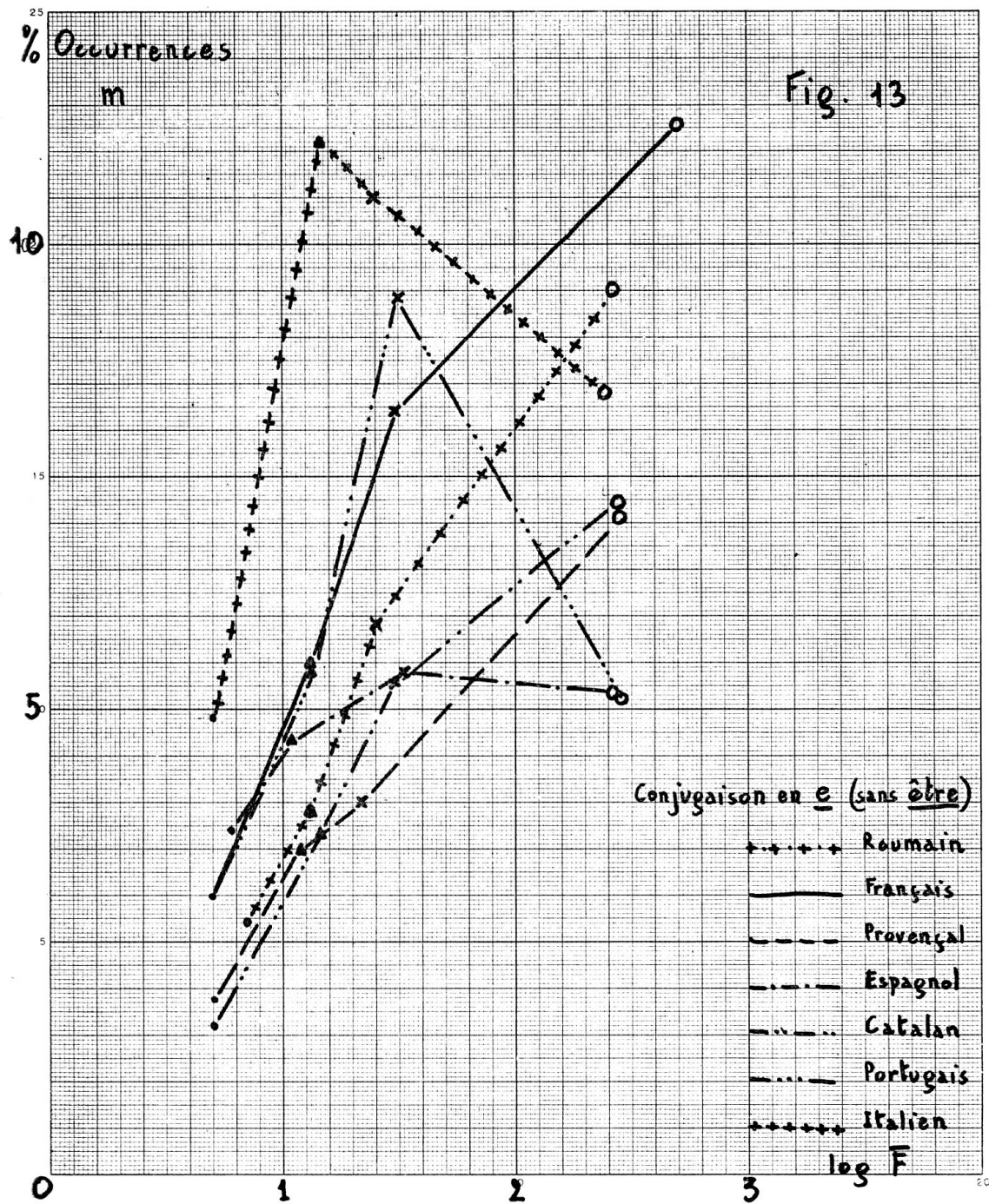
Français 839, Provençal 442, Espagnol 381, Catalan 354, Portugais 413, Italien 324, Roumain 363.

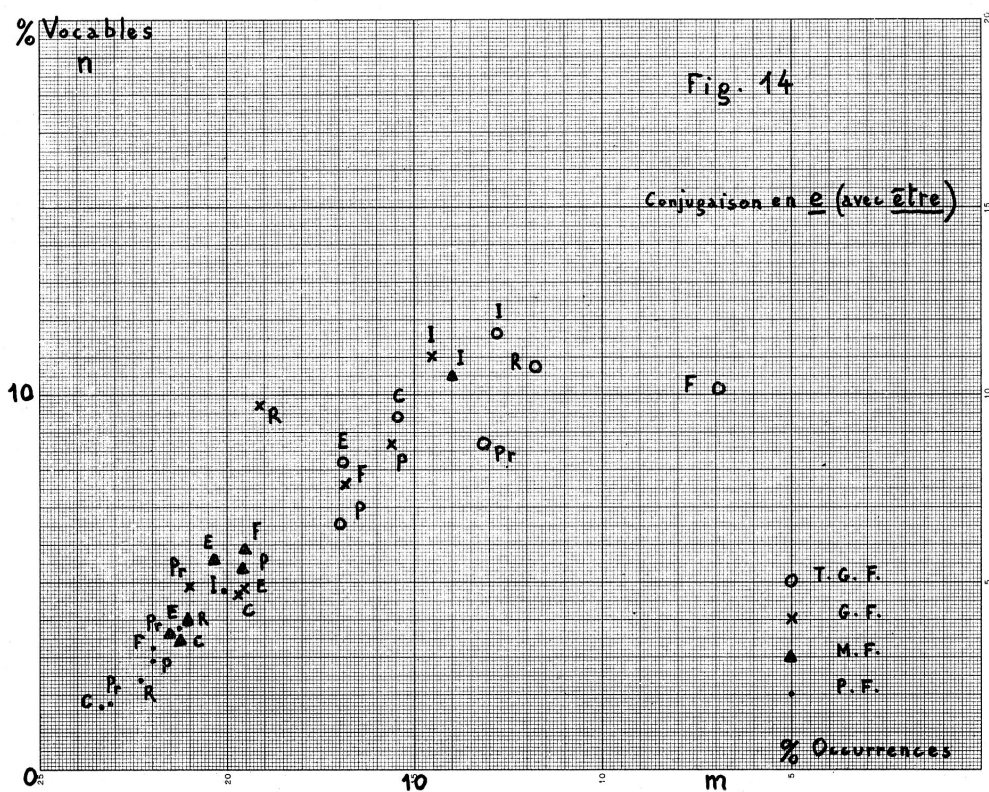
Certes les points représentatifs des graphiques 11 et 14 tendent toujours à se rassembler au voisinage de la bissectrice ; mais alors que, précédemment, les points de la bande de grande fréquence se situaient du côté de l'origine des axes, dans le cas présent, ce sont eux qui s'en écartent le plus. Autrement dit la conjugaison en *e* se caractérise par beaucoup de verbes très employés.

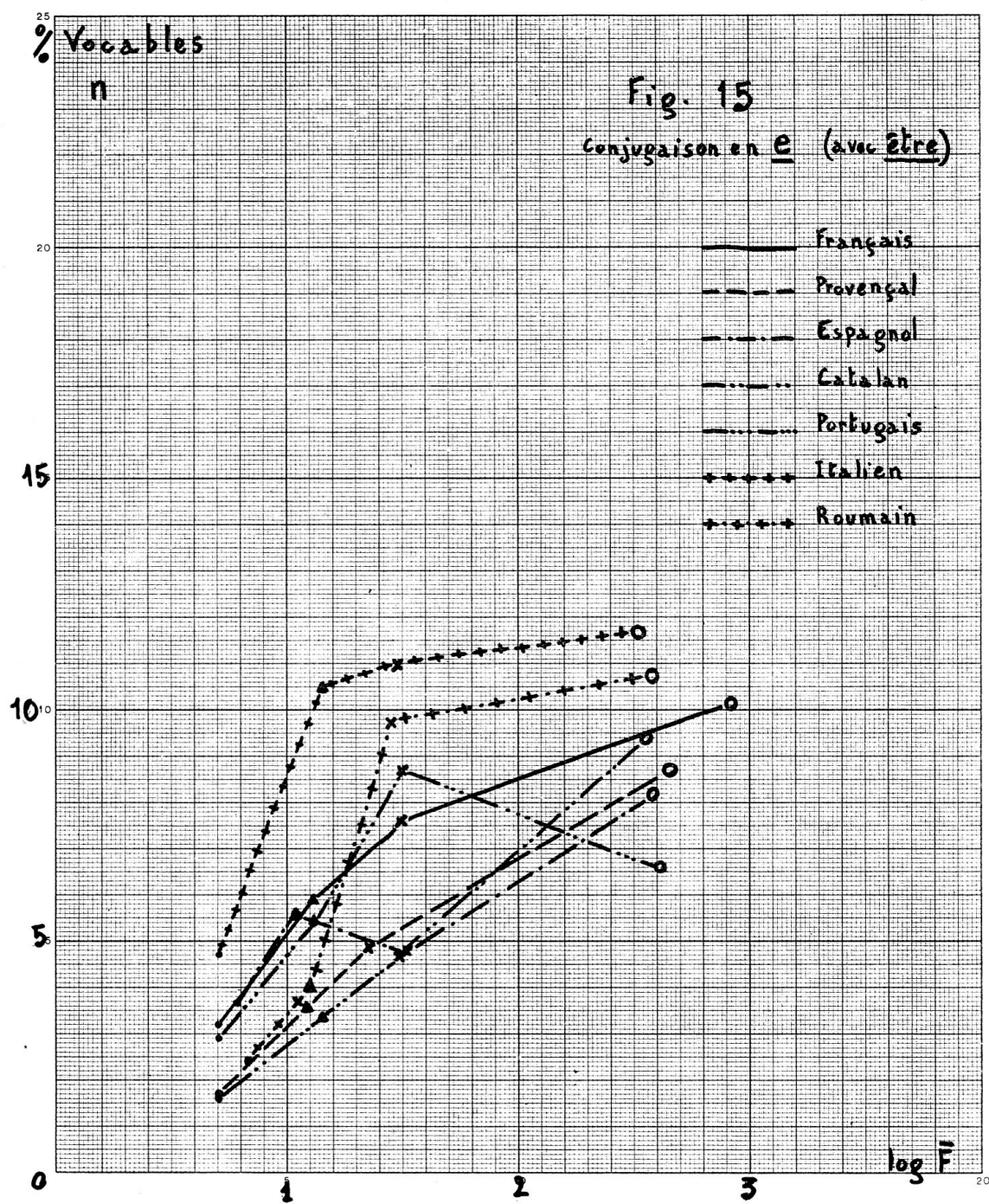
Ces verbes très employés auraient cependant une fréquence moyenne inférieure à la fréquence moyenne de la bande (leurs points sont au-dessus

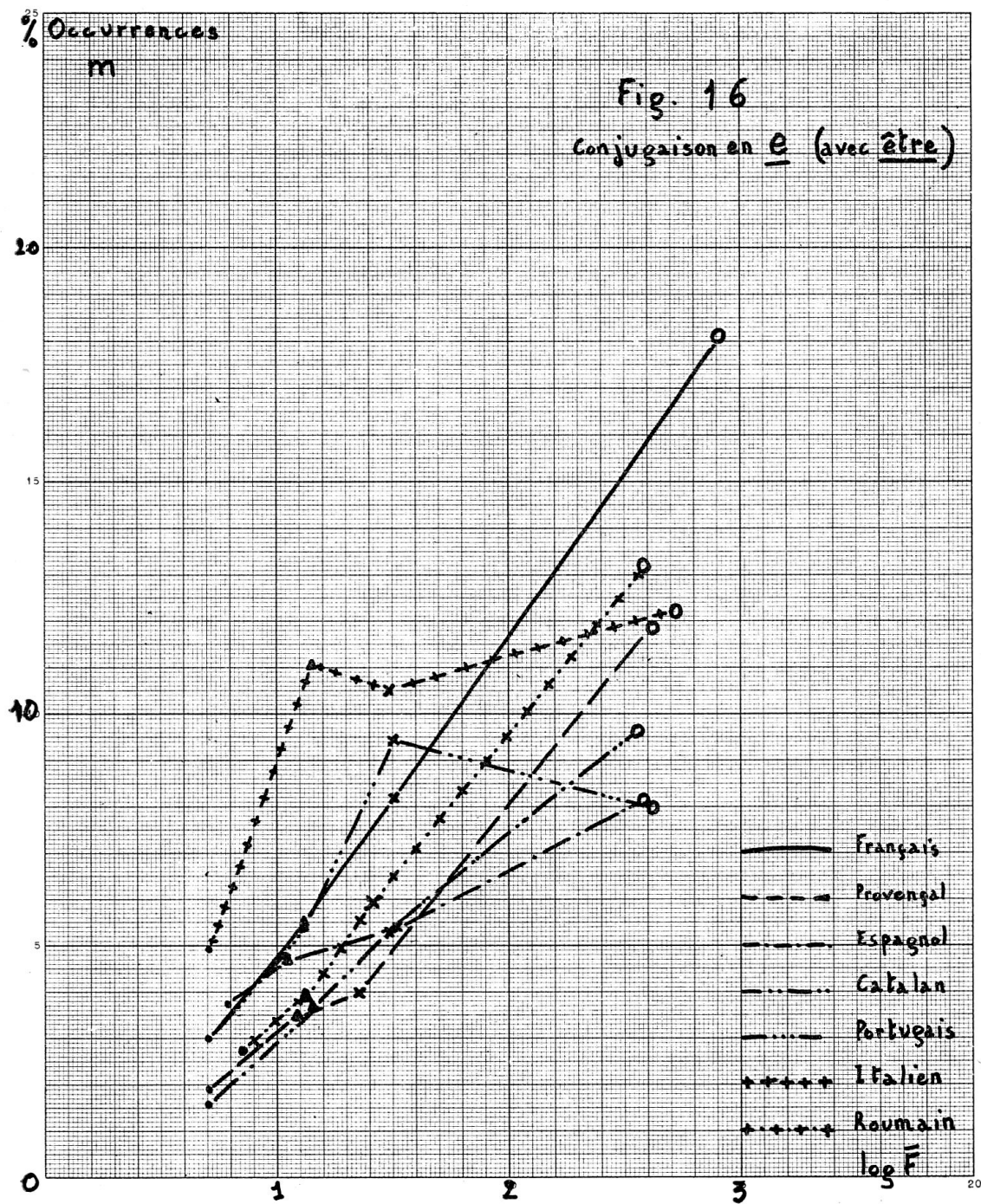












de la bissectrice), mais il en est autrement si l'on fait intervenir le verbe *être*.

Le français se signale par le nombre maximum d'occurrences de T. G. F., l'italien, par le nombre maximum de vocables. Ils sont suivis de très près par le roumain. Plus en arrière viennent le provençal et le catalan ; plus loin encore, l'espagnol et le portugais.

Les courbes représentant les pourcentages de vocables ou d'occurrences en fonction du logarithme de la fréquence, se dessinent avec une allure ascendante absolument différente de l'allure descendante des autres conjugaisons. Les faits sont encore plus nets lorsque le verbe *être* est associé à la conjugaison en *e*.

L'italien a très largement favorisé la conjugaison en *e*, et l'emporte sur toutes les autres langues, sauf pour le pourcentage d'occurrences en T. G. F., où il doit céder au français et au roumain.

Le français offre des courbes continuellement ascendantes, bien au dessus de celle du catalan et du provençal.

Le portugais présente des courbes à maxima très accusés pour la grande fréquence ; il a en effet favorisé la conjugaison en *e*, là où le castillan avait opté pour la conjugaison en *i*.

Quant à l'espagnol, il nous montre des courbes ascendantes mais brisées sauf pour les occurrences avec intervention du verbe *être*.

*
* *

Plusieurs résultats se dégagent donc de cette étude.

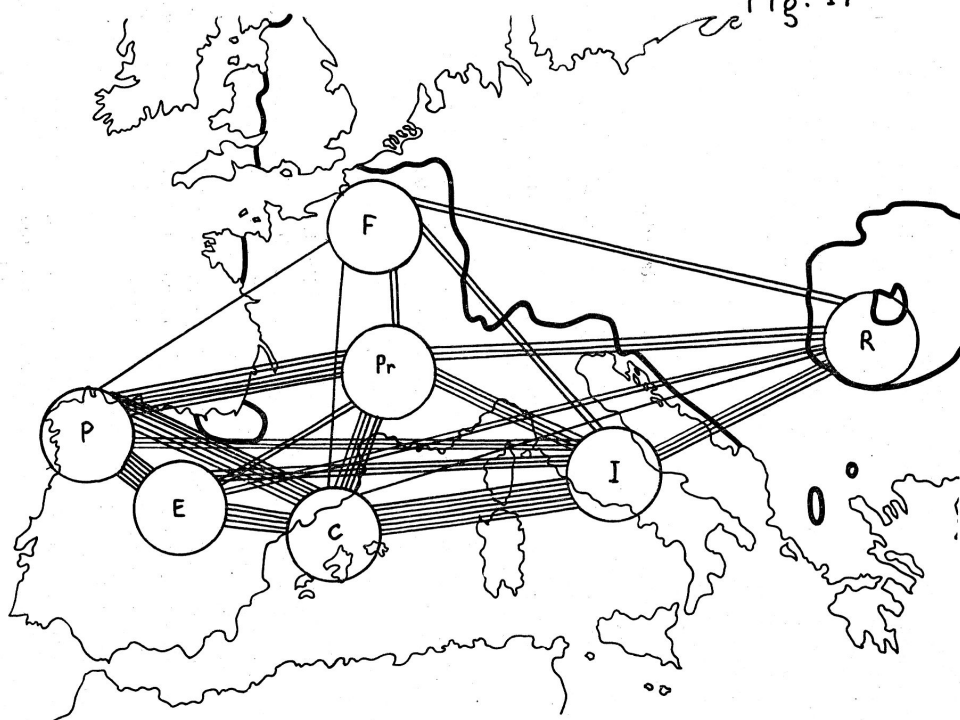
D'abord l'importance de la conjugaison en *e* qui, si elle est inférieure à la conjugaison en *a* par le nombre de vocables, la dépasse largement par le nombre d'occurrences.

Puis l'originalité de la répartition des verbes en *e* à travers les bandes de fréquence. Contrairement aux conjugaisons en *a* et en *i*, la conjugaison en *e* obtient en « très grande fréquence » ses plus forts pourcentages tant de nombre de vocables que d'occurrences ; elle est au contraire peu représentée en « moyenne » et « petite fréquence ».

La conjugaison en *i* ressemble beaucoup à la conjugaison en *a*, mais elle est en gros quatre fois moins importante, sauf en roumain.

L'espagnol a relativement peu de verbes et les emploie peu ; le roumain en a peu qu'il emploie beaucoup ; le français en a au contraire beaucoup et les emploie beaucoup, suivi en cela, bien qu'à distance, par l'italien et le catalan.

Fig. 17



C'est en italien, en roumain et en français que la conjugaison en *e* a maintenu ses plus fortes positions.

Tout au long de l'étude, nous avons été amené à constater des analogies entre telle ou telle langue. Il nous a paru intéressant de rapporter ces observations sur une carte (fig. 17) en symbolisant chaque analogie par un trait rejoignant les deux langues intéressées. Bien entendu, ces diverses analogies n'ont pas toutes la même importance, et il ne faudrait pas donner à cette représentation une signification mathématique rigoureuse.

Avec ces réserves, la carte nous montre un solide réseau de connections entre le portugais, le catalan, le provençal et l'italien, c'est-à-dire les langues les plus conservatrices. L'espagnol est moins solidement associé à cet ensemble, le roumain et le français bien moins encore.

Si nous essayons de caractériser le degré d'« affinité » ou d'« associativité » de chaque langue par le nombre de traits communs avec les autres, nous obtenons le classement suivant : catalan 25, italien et portugais 23, provençal 22, espagnol 19, roumain 12, français 8.

Pour la Romania continue, ce classement des langues, basé sur des données nouvelles, nous montre, une fois de plus, l'isolement du français, avec « *una constitución tan original, tan apartadiza del tipo común al resto, que dentro de la Romania continua el francés resulta inagrupable* » (Amado Alonso, *Miscelánea Fabra*, p. 101), et, à un degré bien moindre, du castillan, « *una radical disidencia lingüística en un rincón de la Cantabria* », selon l'expression de Menéndez Pidal (*El idioma español en sus primeros tiempos*, p. 16).

Henri GUITER.