



## Le volet linguistique du recensement français de 1999 Résultats et analyse appliqués à la Provence plurilingue et au provençal<sup>1</sup>

par Philippe Blanchet, Louis-Jean Calvet,  
Damien Hilléreau & Ewen Wilczyk  
Université de Haute Bretagne,  
Université de Provence (France)

Novembre 2005

### 1. Contexte et méthode de l'enquête

#### 1.1. Contexte général

Lors du recensement quinquennal de la population française en mars 1999, le questionnaire « histoire familiale » (EHF), administré à environ 386 600 personnes (sur environ 60 millions), proposait, parmi vingt-deux groupes de questions, trois questions portant sur certaines pratiques linguistiques des informateurs<sup>2</sup> :

#### TRANSMISSION FAMILIALE DES LANGUES ET DES PARLERS

##### 19. EN QUELLES LANGUES, DIALECTES, OU « PATOIS », VOS PARENTS VOUS PARLAIENT-ILS D'HABITUDE QUAND VOUS ETIEZ ENFANT, VERS L'AGE DE CINQ ANS ?

**VOTRE PERE ou l'homme qui vous a élevé**

**1- vous parlait d'habitude en...**

**2- et aussi en...**

**VOTRE MERE ou la femme qui vous a élevé**

**1- vous parlait d'habitude en...**

**2- et aussi en...**

-Exemples : ALSACIEN, BASQUE, BRETON, CATALAN, CORSE, CREOLE, FLAMAND, GALLO, OCCITAN, PICARD, PLATT, PROVENÇAL, ARABE, ESPAGNOL KABYLE, PORTUGAIS, LANGUES DES SIGNES (LSF)...

**-Pour le FRANÇAIS notez simplement « F »**

-En cas de dialecte ou de « patois » précisez de quel endroit (PICARDIE, BEARN, ROUERGUE, MOSELLE...)

-Pour les langues étrangères, ne notez pas la nationalité mais la langue. Exemple : ne dites pas Algérien, Marocain, Sénégalais, mais ARABE, KABYLE WOLOF, etc.

##### 20. EN QUELLES LANGUES, DIALECTES, OU « PATOIS », PARLIEZ-VOUS A VOS JEUNES ENFANTS, QUAND ILS AVAIENT CINQ ANS (ou leur parlez-vous maintenant s'ils sont plus jeunes) ?

**1- vous leur parliez d'habitude en...**

**2- et aussi en...**

##### 21. ET ACTUELLEMENT, VOUS ARRIVE-T-IL DE DISCUTER AVEC DES PROCHES (conjoint, parents, amis, collègues, commerçants...) DANS D'AUTRES LANGUES QUE LE FRANÇAIS ?

**OUI - NON.**

**Si oui, laquelle ou lesquelles ?**

**1.**

**2.**

<sup>1</sup> Source : INSEE, Enquête études de l'histoire familiale 1999, données mises à la disposition de la DGLFLF par l'INED. Nous emploierons a priori les noms de langues utilisés par l'INED et l'INSEE dans le traitement de ces données, et notamment le nom *provençal* à propos du parler roman local en Provence (qui est le plus courant, cf. *infra*), mais que nous questionnerons au regard des réponses des informateurs. Nos remerciements à J.-M. Eloy et Th. Bulot pour de nombreuses discussions méthodologiques et techniques sur l'exploitation de cette enquête.

<sup>2</sup> Une quatrième question, posée dans le même cadre « Transmission familiale des langues et des parlers », était : « SOUHAITEZ-VOUS RETOURNER VIVRE UN JOUR DANS LA REGION DE VOTRE ENFANCE (que ce soit en France ou ailleurs) ? OUI, NON, vous ne savez pas, vous y vivez déjà.

Cet ensemble de questions était précédé d'une notice explicative en 4 points et schémas, suivis d'informations générales, le dernier des quatre étant celui-ci :

#### 4. En 30 ans, la diversité des langues parlées en famille s'est réduite

	Ce que les adultes parlaient habituellement à leurs enfants vers 1960	Ce que les adultes parlaient habituellement à leurs enfants vers 1990
Le français	84	95
Une autre langue	16	5
	100	100

##### **Les langues parlées en famille**

La majorité de la population française parle français, bien sûr. Mais beaucoup d'entre nous ont été élevés par des parents qui parlaient une langue régionale (Alsacien, basque, breton, catalan, corse, occitan...) ou un parler plus local (béarnais, limousin, picard...). D'autres ont été élevés dans une langue étrangère.

En 1992, une première enquête de l'INSEE a révélé qu'il y a trente ans 16 % des habitants de la France parlaient une autre langue que le français avec leurs enfants. Une génération plus tard, ce chiffre a fortement baissé (tableau 4).

Quelle place le français laissera-t-il aux autres langues ?

Combien y a-t-il de familles bilingues en France ?

Sont-elles en recul ou en progression ?

Dans quelles régions les langues locales se maintiennent-elles le mieux ?

Grâce à vos réponses, on saura pour la première fois comment les divers parlers de toutes origines ont évolué en France depuis 1914.

L'évènement qu'a constitué cette première enquête à échelle nationale sur les langues en France est désormais largement connu des linguistes, voire du grand public<sup>1</sup>. On sait en effet que l'éventualité d'une telle enquête a longtemps soulevé de fortes réticences dans le contexte de l'idéologie du monolinguisme à la française, au point que, aux dires de ses initiateurs, il aura fallu d'abord démontrer, d'une part, qu'elle ne risquait pas de révéler une place trop importante d'autres langues notamment régionales aux côtés du français (d'où la pré-enquête de l'INED publiée par Héran, 1993), et l'organiser, d'autre part, de telle sorte qu'elle mette le français en avant (d'où les modalités de questionnement et les biais qu'elles introduisent, cf. *infra*)<sup>2</sup>. Seules ont eu lieu précédemment quelques enquêtes ponctuelles, en général commanditées par des Conseils régionaux (Languedoc-Roussillon par exemple, cf. Hammel et Gardy, 1994) ou des Conseils généraux (Hautes-Pyrénées par exemple, cf. Dubarry et Dupouts, 1995), par des antennes régionales de l'INSEE (en Bretagne, cf. Laurent, 1992, au Pays basque français en association avec le gouvernement basque espagnol, cf. Oyharçabal, 1999), voire pour des recherches universitaires (sur le breton ou le provençal par exemple, cf. pour un panorama des enquêtes réalisées Broudic 1995 ; Blanchet, 2002). Ces rares enquêtes locales, n'utilisant pas les mêmes protocoles, s'avèrent peu comparables. Il faut en fait remonter à la tentative sommaire et incomplète de V. Duruy (1863, cf. Weber, 1983, 108 et suiv.) pour trouver une enquête nationale sur les langues régionales de France.

Un premier niveau d'exploitation des données linguistiques a permis la publication échelonnée de résultats nationaux synthétiques par les responsables de l'enquête au sein des deux grands organismes d'Etat qui l'ont organisée, l'INSEE et l'INED (Héran, Filhon et Deprez, 2001 ; Clanché, 2002 ; Héran, 2003 et 2004). Un second niveau d'exploitation a permis la publication en cours de résultats régionaux globaux, soit par des agences régionales de l'INSEE ayant réalisé les analyses (Duée, 2002 ; Deguillaume et Amrane, 2002 ; Le Boëtté, 2003), soit dans des revues régionales de l'INSEE par des centres de recherche en sociolinguistique (Blot, Eloy et Rouault, 2004)<sup>3</sup>.

<sup>1</sup> Ce n'était d'ailleurs pas la seule ni la moindre des innovations, puisque pour la première fois depuis sa création en 1954, cette enquête a inclus les hommes dans son échantillonnage.

<sup>2</sup> Le fait que les départements et territoires d'outre-mer ne soient pas pris en compte, tout en obéissant aussi à d'autres motivations, exclut de fait les zones de France où les langues « locales » sont les plus utilisées et où le français est le plus fortement concurrencé.

<sup>3</sup> Cet article était rédigé lorsqu'est paru le n° 112 de *Langage et Société* (juin 2005) contenant les exploitations de l'EHF pour la Corse par G. Moracchini (« Aspects de la situation sociolinguistique de la Corse.

## 1.2. Les projets d'exploitation sociolinguistique

Un troisième niveau d'exploitation a été mis en place par une convention entre l'INED et le Ministère de la Culture, par une convention signée en 2003 qui autorise le DGLFLF (service de Ministère de la Culture) à confier à des centres de recherche en sociolinguistique les données issues du recensement afin d'en réaliser une exploitation plus fine et plus approfondie d'un point de vue sociolinguistique. Six projets en ce sens ont été suscités et confiés à 7 équipes universitaires : deux projets à Montpellier III-Paul Valéry (laboratoires DIPRALANG EA 739 et UMR 5475 REDOC, un à Strasbourg III-Marc Bloch (département d'études allemandes), un à Corte - Université de Corse (UMR 6039 BDLC), un à Amiens-Jules Vernes (laboratoire LESCLAP)<sup>1</sup>, et enfin un à Aix-Marseille I-Université de Provence (Institut de la Francophonie) et Rennes 2-Université de Haute Bretagne (CREDILIF EA 3207)).

### 1.2.1. Le projet de l'Institut de la Francophonie et du CREDILIF

Le projet soumis conjointement par ces deux derniers centres de recherche, sous la responsabilité de L.-J. Calvet et de Ph. Blanchet porte sur deux volets complémentaires :

- Les noms des langues comme indices des représentations sociolinguistiques des informateurs et des populations dans l'ensemble de la France métropolitaine.
- Les langues en région Provence-Alpes-Côte d'Azur.

Il est vite apparu qu'il nous était à la fois matériellement impossible de travailler sur l'ensemble des données (plus de 380.000 informateurs, dont plus de 65.500 ayant mentionné près de 6700 pratiques linguistiques autres que le français catégorisées en près de 400 « langues » ou « dialectes » dans la base de données...) et néanmoins nécessaire de comparer des ensembles différents (langues différentes, régions différentes, etc.) afin de déceler d'éventuels biais et de valider des tendances générales ou spécifiques. Nous avons donc choisi de travailler sur trois aires cohérentes sur les plans linguistique et historique et à propos desquelles nos équipes ont des compétences précises :

- la Bretagne historique (région administrative actuelle + département de Loire-Atlantique),
- la Normandie (les deux régions administratives de Haute et Basse Normandie),
- la Provence (la région administrative Provence-Alpes-Côte d'Azur à laquelle est retranché le département des Alpes-Maritimes).

Cet échelon a été retenu parce que c'est le plus petit niveau de représentativité statistique de l'échantillon, construit par région. En deçà (au niveau du département, voire du canton ou de la commune), la représentativité n'est pas garantie d'un point de vue statistique (certains départements sont totalement absents ou représentés uniquement sur certaines variables, par exemple uniquement des femmes — ce n'est heureusement pas le cas pour le seul département dont nous avons modifié l'inscription régionale, la Loire-Atlantique)<sup>2</sup>. Une demande d'autorisation d'avoir accès à des critères de zonage plus précis, par exemple pour étudier la situation de Marseille, pertinente en sociolinguistique urbaine, est d'ailleurs restée sans réponse<sup>3</sup>. Les situations sociolinguistiques urbaines engagent pourtant de nos jours des enjeux cruciaux qui ont ouvert des dynamiques nouvelles dans la recherche en sociolinguistique : effet de l'urbanisation sur les pratiques linguistiques, analyse des situations sociales urbaines à partir des indicateurs linguistiques (cf. Calvet, 1994, Bulot et alii, 2001, Bulot et Messaoudi, 2003). En outre, de nombreux travaux avaient préalablement montré qu'une certaine conscience linguistique — voire identitaire — régionale existe de façon significative dans les régions ainsi délimitées :

- du point de vue des « langues régionales », les 2/3 de la Loire-Atlantique situés au nord de la Loire sont très clairement couverts par une conscience identitaire bretonne et par des parlers locaux de type « gallo » (et breton celtique sur la zone côtière)

---

Les apports de l'enquête famille », qui émet des réserves méthodologiques proches des nôtres) et par J.-M. Géa (« Immigration et contacts de langues en Corse »).

<sup>1</sup> C'est de ce projet qu'est issu la première publication de Blot, Eloy et Rouault, 2004.

<sup>2</sup> Nous reviendrons sur ce point *infra*.

<sup>3</sup> Même si un certain nombre de critères présents dans la base permettent d'identifier indirectement les informateurs résidant en zones urbaines et rurales, et souvent la capitale régionale du fait de la « tranche d'unité urbaine » signalée (par rapport au nombre d'habitants). Cf. *infra*.

comme le reste de la Bretagne historique ; les Alpes-Maritimes sont constituées au 3/4 par l'ancien Comté de Nice avec une conscience linguistique morcelée et complexe (les parlers locaux y sont massivement identifiés de façon très localisée, cf. Dalbera, 1994) et une conscience linguistique majoritairement non provençale, alors que le reste de la région administrative recouvre l'ancienne Provence historique, à l'exception partielle du département des Hautes-Alpes qui a, dès lors, fait l'objet d'un traitement spécifique (rendu de toute façon nécessaire par un problème de codage de la base INED, cf. *infra*). Avec la Bretagne et la Provence, nous avons affaire à deux régions aux « langues régionales » bien repérées (le breton et le provençal) ; avec la Normandie et la Bretagne, nous avons affaire à deux zones d'oïl contigües (dites « gallo » et « normande »).

- du point de vue des « langues étrangères et de l'immigration<sup>1</sup> », nous avons affaire à des régions accueillant toutes des personnes porteuses de ces langues, mais à des degrés fort divers (la Provence a accueilli beaucoup plus de migrants que la Bretagne), et divers travaux ont montré que l'intégration est liée à un sentiment d'appartenance local ou régional d'autant plus intégrateur qu'il est marqué (cf. par ex. Césari *et alii*, 2001 ; Blanchet, 2003a).

La comparaison (qui nécessite à la fois semblance et dissemblance) est donc possible à un niveau régional affiné.

N'est proposée ici qu'une partie des résultats de ce programme, ceux portant sur le provençal dans le contexte plurilingue de la Provence.

### 1.2.2. Cadre méthodologique

Les deux agences régionales de l'INSEE, en Provence et en Bretagne, ont été sollicitées pour une collaboration au programme de travail. L'INSEE Provence ne s'y est pas intéressée (et n'a d'ailleurs produit à ce jour aucune analyse de l'enquête à son niveau régional). À l'INSEE Bretagne, des personnels s'étaient déclarés intéressés mais la direction a mis son veto à toute participation à ce projet (elle préparait d'ailleurs sa propre exploitation régionale des données, publiée par Le Boëtté en 2003).

L'exploitation des données a donc été réalisée conjointement par des statisticiens de Rennes 2 et des sociolinguistes d'Aix et de Rennes 2. Damien Hilléreau et Ewen Wilczyk, diplômés en mathématiques appliquées aux sciences sociales (MASS) de Rennes 2, ont travaillé pendant deux à trois mois en 2003 au CREDILIF sous la co-direction de Ph. Blanchet et d'A. Mom, maître de conférences de mathématiques au dpt. MASS de Rennes 2. La méthode employée a donc consisté à entrer dans l'analyse de la base par deux démarches complémentaires et articulées l'une à l'autre. La définition des régions a été fondée sur des critères sociolinguistiques (cf. *supra*). Après avoir extrait de la base nationale les trois sous bases constituées par les informateurs résidant dans les trois régions concernées, les statisticiens se sont livrés à une analyse mathématique conforme aux principes de la statistique. Pour résumer, il s'agit de faire « parler les chiffres » en appliquant une analyse factorielle des correspondances multiples (AFCM) sans hypothèse sociolinguistique préalable. Dans un deuxième temps, les résultats de l'analyse factorielle ont fait l'objet d'une interprétation sociolinguistique induisant de nouvelles interrogations de la base. Enfin, les données ont analysées, selon une méthode inverse, à partir de questionnements sociolinguistiques.

### 1.2.3. Méthode d'analyse statistique

Cette partie de l'analyse permet de voir les relations qui lient toutes les variables entre elles et de les placer sur un axe factoriel. L'AFCM se propose d'étudier les relations qui lient Q variables (totalisant J modalités) entre elles. Les données se présentent soit sous la forme d'un tableau disjonctif complet Z, soit sous la forme d'un tableau B de Burt. Elle consiste à effectuer une analyse factorielle simple sur un tableau disjonctif complet ou sur un tableau de Burt.

Le *tableau disjonctif complet* Z est une matrice n x J qui a autant de lignes que d'individus (n) et autant de colonnes que de modalités (J) :

$$Z_{n \times J} = [Z_{ij}]$$
$$Z_{ij} = 1 \text{ si l'individu } i \text{ prend la modalité } j, 0 \text{ sinon.}$$

<sup>1</sup> Non distinguée et peu distinguables dans l'enquête malgré l'un des recodages de l'INED, à l'exception possible d'une forte proportion d'informateurs déclarant l'anglais.

Le *tableau de Burt* B est une matrice J x J composée de Q tableaux de contingence croisant deux variables (q et q'). B = Z<sup>t</sup>Z. À l'intersection de la i<sup>ème</sup> ligne et de la j<sup>ème</sup> colonne se trouve le terme B<sub>ij</sub> qui est le nombre d'individus prenant simultanément les modalités i de la variable q et j de la variable q'.

L'interprétation des résultats est réalisée de la façon suivante :

Le nombre de *valeurs propres* non triviales  $\lambda_\alpha$  est J-Q. Lorsque l'on effectue une AFCM sur un tableau disjonctif complet, les pourcentages d'inertie expliqués par les axes sont faibles. En effet, ce ne sont que des mesures pessimistes de l'information réellement restituée. Les mêmes ratios effectués sur un tableau de Burt donnent de meilleurs résultats mais peuvent tout de même être améliorés.

Ceci provient du fait que l'AFCM analyse un tableau de Burt, c'est-à-dire simultanément les blocs diagonaux (croisant une question avec elle-même) qui ont la plus grande inertie possible mais qui ne portent pas d'information et les blocs extra-diagonaux qui contiennent toute l'information importante mais qui ont une inertie bien plus faible. En effet, on montre que les tableaux de contingence carrés et diagonaux (comme les blocs diagonaux du tableau de Burt) ont une inertie maximale. Le calcul des taux modifiés de Greenacre permet de calculer les pourcentages d'inertie expliqués par les axes en tenant compte de cette remarque.

Enfin notons que les graphiques utilisés sont pseudo-barycentriques. Autrement dit, il existe, suivant les axes un facteur de dilatation qui n'est pas le même. Ce facteur est égal à l'inverse de la racine carrée de la valeur propre. De ce fait, comme la deuxième valeur propre est plus petite que la première, les distances entre les modalités des deux variables ne sont plus directement interprétables. De plus, le phénomène étudié est plus que bi-dimensionnel ; ainsi deux variables proches l'une de l'autre dans le repère formé des deux premiers axes factoriels, peuvent se révéler finalement très éloignées selon la direction du troisième axe. Ainsi le calcul des Cos<sup>2</sup> est très important pour donner du sens à ce graphique.

Un test de significativité d'une modalité supplémentaire (Test de Lebart-Morineau) a été appliqué. Une variable supplémentaire est une variable que l'on n'intègre pas comme les autres dans l'analyse. Elle ne contribue pas à créer les axes. En revanche on la projette sur les axes obtenus pour voir où elle se place. On peut décider de mettre une variable (et ses modalités) en supplémentaire si elle est trop importante lors d'une étude précédente, pour conforter l'interprétation des axes ou pour expliquer ces variables à l'aide des variables actives (=non mises en supplémentaires).

Pour une variable supplémentaire, nous pouvons calculer la coordonnée d'une de ses modalités. Soit s la modalité supplémentaire. On calcule son profil conditionnel, puis sa coordonnée sur l'axe a Csa. Il est utile de déterminer si cette modalité supplémentaire est « liée » avec un axe a donné. Bien entendu une modalité sans lien aura une coordonnée « faible ». Afin de quantifier cette notion de faible nous avons utilisé le test de Lebart-Morineau :

La modalité s est prise par n<sub>s</sub> individus parmi les n. L'hypothèse nulle à tester est alors : Ho : « Les n<sub>s</sub> individus sont choisis au hasard parmi les n. On est donc en présence d'un plan de sondage à Probabilités Egales Sans Remise (PESR). » En effet si les n<sub>s</sub> individus sont aléatoirement répartis alors ils ne dépendent pas de l'axe.

D'après Lebart et Morineau on obtient la statistique S suivante :

$$S = \sqrt{\frac{n-1}{n-n_s}} \sqrt{n_s} c_{sa} \sim N(0,1)$$

On rejette donc Ho si |S| > 2 pour un risque de première espèce de 5 %.

Si ce lien s'avère significatif nous pouvons grâce aux interprétations des axes créés par les variables initiales (actives) relier cette modalité supplémentaire aux modalités initiales. On peut donc expliquer cette variable supplémentaire à l'aide des variables actives.

#### 1.2.4. Informations sociolinguistiques recherchées

*Données brutes :*

- Liste et % des langues citées dans chaque région
- % de personnes déclarant avoir reçu la langue régionale locale (LRL) ou une langue de l'immigration (LI)

- % de personnes déclarant avoir transmis la LRL ou LI
- % de personnes déclarant utiliser la LRL ou LI
- % d'emplois des différents noms spontanés de la LRL ou LI
- % de personnes mentionnant une autre LR que la LRL
- % de locuteurs déclarés de la LRL et d'une langue de l'immigration (LI)
- Liste et % des LI mentionnées aux côtés de la LRL (et inversement)
- % de locuteurs déclarés de la LRL ou LI comme 1ère langue
- % de locuteurs déclarés de la LRL ou LI dans les métropoles urbaines régionales et départementales (Rennes, Brest, St Brieuc, Vannes, Nantes, Rouen, Le Havre, Dieppe, Yvetot, Caen, Alençon, Marseille, Aix, Avignon, Toulon, Digne, Gap)

*Croisement de ces données :*

- Entre elles.
- Avec le facteur *sexe*
- Avec le facteur *année de naissance*
- Avec le facteur « *département* »
- Avec le facteur *rural / urbain*
- Avec le facteur *CSP*
- Avec le facteur *niveau d'études*

## **2. Problèmes méthodologiques, biais, limites et pertinence de l'enquête EHF**

### **2.1. Contraintes méthodologiques pour une situation diglossique**

La situation sociolinguistique française étant caractérisée par les effets d'une diglossie marquée, toute enquête y est difficile (cf. Eloy, 1998 pour un panorama), et d'autant plus difficile si elle est quantitative, puisque certaines pratiques linguistiques y sont réservées à des situations qui échappent à la présence du chercheur, puisque leur stigmatisation en fait des objets « cachés », puisque la présence même du chercheur (a fortiori en situation explicite d'enquête institutionnelle) oriente les pratiques (et les discours sur les pratiques), notamment en direction des pratiques et des représentations « légitimes » valorisantes.

Cela nécessite donc des protocoles adaptés, soigneusement mis au point afin d'en réduire les biais. Il est notamment plus efficace de procéder selon une progression méthodologique allant de l'« analogique », étude qualitative des pratiques et représentations linguistiques en contexte social, marqués par la continuité, au « digital », étude analytique d'éléments construits comme discontinus (Calvet, 2004, p. 31 et suiv., notamment pp. 52-55). Ou, pour le dire en d'autres termes, selon une « méthode en sablier » qui va du global, observation participante en contexte, à l'analytique, enquêtes explicites y compris quantitatives, pour revenir au global, c'est-à-dire à des synthèses interprétatives (Blanchet, 2000, 40 et suiv.). Au fond, en effet, le quantitatif est une modalité du qualitatif, puisque les questions et les catégories selon lesquelles les observables quantitatifs sont construits sont des éléments signifiants, procédant d'interprétations préalables (y compris sous la forme d'hypothèses elles-mêmes nécessairement élaborées selon des interprétations intuitives ou raisonnées de l'expérience subjective)<sup>1</sup>.

L'apparente objectivité des chiffres est trompeuse : ils sont nécessairement issus de procédures de mesure subjectives et n'ont de valeur que par l'interprétation contextualisée qui en oriente la « lecture » vers certaines significations (cf. Blanchet, 1995, pour une analyse critique d'enquêtes quantitatives sur le breton). Les « données » sont toujours des *observables* construits par l'enquêteur.

L'une des limites fréquentes des enquêtes sociolinguistiques est, en revanche et d'un certain point de vue, leur manque de confirmation des observables « significatifs » sur le plan qualitatif par des « mesures » quantitatives validées comme « représentatives » sur un plan statistique. Leurs échantillons sont souvent trop réduits d'un point de vue statistique<sup>2</sup>, quoique significatifs en termes qualitatifs par la méthode de « saturation » du corpus. De ce point de vue, l'EHF apporte des éléments complémentaires utiles à nos recherches, sa validité statistique étant garantie.

<sup>1</sup> Cette analyse des critères quantitatifs doit beaucoup à des échanges avec Didier de Robillard.

<sup>2</sup> Même si des méthodes de statistiques sur petits échantillons existent (Cf. Marien et Beaud, 2003).

## 2.2. Biais et limites de l'EHF

### 2.2.1. Problèmes du questionnaire et de sa passation

L'enquête EHF de 1999 n'a pas été élaborée selon cette méthode inductive à partir du terrain, malgré la recherche d'une « solution qui reconnaît pleinement l'existence des phénomènes de diglossie » (Héran, 2004, p. 14). Elle présuppose des catégorisations et des pratiques sans tenir suffisamment compte du terrain, du contexte social, du poids de l'idéologie monolingue française. Il ne s'agit pas ici de minimiser les efforts grâce auxquels cette première enquête nationale a pu avoir lieu, mais, tout en comprenant les limites matérielles imposées à l'exercice (notamment l'espace octroyé aux questions sur le document d'enquête), d'en analyser les biais (propres à toute enquête), les limites (idem), pour en proposer l'exploitation la plus prudente et la plus rigoureuse possible, aussi bien d'un point de vue scientifique qu'en termes d'interventions médiatiques et glottopolitiques (cf. Eloy, 1998). Lors de la parution des premiers résultats nationaux du volet linguistique l'enquête EHF 1999, il en a été rendu compte en soulevant certaines des principales difficultés et en pointant certains biais de l'enquête (Blanchet, 2003b) :

- renforcement de la pression diglossique par le document de présentation de l'enquête et par les consignes de passation (cf. ci-dessus), en insistant sur la prééminence du français (« *la majorité de la population française parle français, bien sûr* » ; *pour le français, notez simplement F<sup>1</sup>*) et sur une évolution sociale d'abandon des autres langues au profit du français (« *en 30 ans, la diversité des langues parlées en famille s'est réduite* ») ;
- ambiguïté sur les notions utilisées : que signifie exactement *parler* une langue (questions 19 et 20), *d'habitude* (question 19 — on notera que l'absence de cet élément à la question 20 peut être compris comme signifiant soit « très occasionnellement » soit, à l'inverse « systématiquement »), *discuter, proches, parents* (question 21) ? Des interprétations variables ont dû être à l'œuvre, notamment des représentations très répandues selon lesquelles *parler une langue voir être bilingue* signifie « avoir une maîtrise parfaite de cette langue » ;
- induction et ambiguïté des catégorisations linguistiques proposées : *béarnais, rouergat* sont des exemples de « dialectes » ou de « patois », *l'algérien* ou le *marocain* sont forcément de l'« arabe », bien que ces dénominations soient de fait utilisées dans d'autres systèmes de catégorisations dans les sociétés concernées. La présentation de l'enquête par l'INSEE (dossier interne d'accompagnement des données, p. 3-17) publiée par Cassan et alii (2000, 30) déclare d'ailleurs explicitement : « *Le questionnaire met délibérément sur le même plan les langues régionales et les langues nationales, car ces dernières sont souvent des dialectes, tandis que les premières peuvent être des prolongements frontaliers de langues nationales* », ce qui révèle des catégorisations a priori distinctes de celles émanant des pratiques sociales ; de fait, les termes utilisés par les informateurs ne recouvrent pas nécessairement les pratiques linguistiques auxquelles on pourrait s'attendre : beaucoup de locuteurs des « langues d'oïl » considèrent qu'ils parlent français en les parlant, le terme *patois* employé sans aucune précision par une quantité importante d'informateurs reste alors très flou.
- hypothèse selon laquelle la transmission parentale avant l'âge de scolarisation obligatoire<sup>2</sup> est un indicateur clé des pratiques linguistiques familiales et de leur transmission, alors que de nombreuses études de terrain ont montré que d'autres facteurs de pratiques et de transmission existent et ont été probablement au moins aussi importants au cours du XXe siècle — y compris sur le plan quantitatif : rôle des grands-parents, des sociabilités de pairs, des milieux professionnels et de loisirs, des rituels festifs, etc. ;
- sur-représentation de zones hyperurbaines et sous-représentation de zones semi-rurales et rurales : en Provence par exemple, 82,4 % des informateurs de l'EHF pour les Bouches-du-Rhône résident dans l'agglomération marseillaise qui ne représente que 67,94 % des habitants du département et 60,67 % des informateurs varois résident dans l'agglomération toulonnaise qui ne représente que 49,49 % des habitants du département.

<sup>1</sup> Cette phrase en gras.

<sup>2</sup> Âge théorique : la totalité des enfants est aujourd'hui scolarisé à partir de 4 ans, et l'accès aux classes maternelles s'est massivement développé dès les années 1960.

On a donc tout lieu de supposer que ces modalités d'enquêtes ont induit une sous-déclaration des pratiques linguistiques visées autres que celles du français<sup>1</sup>, et, assurément, n'envisagent que certains types de transmission et de pratiques parmi d'autres probablement aussi importants. De fait, on observe au moins deux indices de sous-déclaration :

- « *L'enquête famille restant facultative, 20 % des personnes sollicitées n'ont pas rempli ce questionnaire* » (Héran, 2004, 11).
- On constate des écarts entre les déclarations des parents et des enfants (les parents déclarant en général moins transmettre que les enfants ne déclarent avoir reçu), cf. Le Boëtté, 2003, 22 et *infra*).
- La plupart des enquêtes quantitatives réalisées selon des protocoles sociolinguistiques plus affinés donnent des chiffres plus élevés (de 30 % à 200 %), notamment en ce qui concerne la pratique actuelle déclarée...

### 2.2.2. Problèmes dans l'interprétation et l'exploitation des résultats

Il en va de même, du reste, pour l'interprétation et l'exploitation des résultats. Une simple comparaison des résultats publiés révèle des différences frappantes.

Ainsi, pour Héran *et alii* (2001), la question 19 permet de dire que « *les adultes interrogés sont nombreux à se souvenir<sup>2</sup> que leurs parents leur parlaient, associée ou non au français, une autre langue* ». La question 19-2 (*et aussi en...*) est interprétée comme révélant la transmission *occasionnelle*, par opposition à la transmission *habituelle* déclarée au 19-1, et l'on en tire la conclusion « *Le contraste de ces deux types de pratiques est éclairant* », alors que rien ne dit que le *et aussi* de la question 19-2 exclut automatiquement la notion « d'habitude » pour les répondants. Les résultats présentés de la question 20 ne mentionnent plus du tout l'âge stipulé dans la question et sont donc étendus aux enfants à tous âges. Enfin, les résultats de la question 21, qui n'est pas formulée en ces termes, sont présentés comme indiquant « *les langues pratiquées en dehors de la relation aux enfants* » et les chiffres rapportés avec la précaution « *20 % des adultes (...) disent que...* ».

L'autre exploitation globale au niveau national (Clanché, 2002) est moins mesurée dans ses formulations : le verbe *déclarer* ou modalisation similaire y est très rare, la notion de transmission *occasionnelle* est également employée comme opératoire, et les résultats de la question 21 sont généralisés en « *Proportion d'utilisateurs actuels d'autres langues que la français* », la précision *avec des proches* étant éliminée.

Dans Héran (2004, pp. 15-16) on emploie également le verbe *se souvenir* pour rendre compte des réponses à la question 19, mais on y passe très vite à des généralisations moins modalisées : « *une fois sur deux, cette transmission a eu lieu...* » (p. 16) ou « *le taux de retransmission global d'une génération à l'autre varie très fortement selon les langues...* » (p. 19), ou encore, en conclusion (p. 24), « *au total, l'enquête Famille permet de retracer de façon dynamique les processus de retransmission et d'intégration linguistique des divers courants migratoires* » par exemple.

On retrouve la même généralisation observée en conclusion dans Héran (2003, 6) : « *Le taux de retransmission aux enfants varie entre 90 % et 9 % selon les langues* » et les mêmes glissements dans les analyses régionales publiées par l'INSEE. La plus prudente est l'analyse picarde, menée par des universitaires, qui emploient très régulièrement (mais pas systématiquement) le verbe *déclarer* pour attirer l'attention du lecteur sur l'écart qu'il peut y avoir entre pratiques déclarées et pratiques effectives, et qui précisent dans leur conclusion à propos de l'enquête « *son caractère quantitatif limite les interprétations possibles* » (Blot *et alii*, 2004, p. 4). Le Boëtté, qui élargit et relativise le sujet de l'enquête, signale en conclusion « *il existe toutefois un biais déclaratif* » (2003, p. 22). La publication aquitaine généralise — abusivement — la portée de la question 19 en [x % des informateurs] « *disent avoir entendu parler occitan pendant leur petite enfance* » (par qui que ce soit, donc, et pas seulement par leurs parents s'adressant à eux).

<sup>1</sup> Même celles de l'anglais, seule autre langue socialement très valorisée en France, ne sont pas sûres : outre les effets de la définition perfectionniste courante de *parler une langue*, il faut compter avec les effets du nationalisme linguistique français, notamment institutionnel, qui tend à stigmatiser pour certains aspects la pratique de l'anglais perçue comme concurrente avec celle du français.

<sup>2</sup> On notera la précaution justifiée de formulation.



### 2.2.3. L'exemple des dénominations des pratiques linguistiques

L'un des éléments importants que cette enquête peut apporter est constitué par les représentations sociolinguistiques impliquées par les dénominations des pratiques, et notamment les catégorisations ou classifications linguistiques fonctionnelles. Ces informations figurent dans la base de données (y compris les fautes de frappe manifestes à la saisie). Elles ont par ailleurs fait l'objet de regroupements par recodage :

- sur la base de la classification internationale des langues de l'*Ethnologue*, modifiée et complétée (382 langues ou « modalités »),
- par regroupement de ces langues en 28 groupes de langues ou « modalités » :

**Figure 1 : Recodage EHF en 28 groupes linguistiques**

Langues	Codage
français	00
alsacien	11
breton	12
catalan	13
platt mosellan	14
basque	15
corse	16
francontois, franco provençal	17
flamand	18
langues d'oc	19
langues d'oïl	20
créoles à base française	21
portugais	41
espagnol	42
italien	43
allemand	44
polonais	45
anglais	46
serbo-croate	47
langue ou dialecte italien	48
autres langues d'Europe	49
arabe	51
langues berbères	52
autres langues d'Afrique	53
turc	54
kmer, lao, thaï, vietnamien	55
langues de Chine et du Tibet	57
autres langues	60

Trois autres regroupements en 40, 14 ou 4 modalités sont également disponibles, le tout couplé à des informations sur le degré de fiabilité du recodage, la zone géographique d'origine de la langue, etc.

Ces regroupements et recodages sont loin d'être exempts d'approximations, de choix discutables, voire d'erreurs manifestes, tout cela à des degrés divers selon les pratiques linguistiques déclarées ou envisagées. Selon l'INED, 94,7 % des déclarations des informateurs ne posent pas de problème de recodage selon la classification de l'*Ethnologue*. Le tableau ci-dessus, qui a été utilisé par l'INED pour produire un certain nombre de pourcentages (tableau dit « variables L-SIMPL<sup>1</sup>»), a néanmoins dû être recomposé par nos soins à partir des fichiers informatique de la base de données, car il figure dans les documents d'accompagnements sous la forme :

**Figure 2 : Codes erronés**

francontois, franco provençal	17
flamand, langues d'oc	18
langues d'oïl	19
franco provençal	20

<sup>1</sup> Le « flamand langues d'Oc » y obtient même 0,2 % des déclarations de l'ensemble des informateurs de l'EHF.

Or il apparaît, au-delà de la coquille informatique, que le groupe linguistique construit par certains linguistes sous la dénomination *franco-provençal* (ici code 17) pose effectivement des problèmes dans cette enquête, puisqu'y sont regroupées par exemple les déclarations suivantes et leurs codages EHF d'après *The Ethnologue* :

**Figure 3** : Recodage sous le regroupement 17

patois	fra (franco-provençal)
patois_ardèche	lpn (provençal alpin)
patois_dauphinois	lpn (provençal alpin)
patois_vosgien	fcc (franc-comtois)
patois_savoyard	fra (franco-provençal)
dialecte_suisse	sch (langue ou dialecte suisse) <sup>1</sup>
dialecte_suisse	sch (langue ou dialecte suisse)
patois	fra (franco-provençal)
patois	lpn (provençal alpin)
patois	lpn (provençal alpin)
patois_du_pays	fra (franco-provençal)
patois_du_pays	fra (franco-provençal)
comtois	fcc (franc-comtois)
comtois	fcc (franc-comtois)
vosgien	fcc (franc-comtois)
patois_vosgien	fcc (franc-comtois)
patois_vosgien	fcc (franc-comtois)

De fait, en ce qui concerne le département des Hautes-Alpes qui entre dans la région que nous nous sommes donnée pour objet d'étude, nous avons très vite découvert que les déclarations, toutes sous la dénomination *patois*, avaient été recodées en franco-provençal, alors que, à la fois pour des raisons linguistiques et sociolinguistiques (conscience linguistique des populations), il s'agit plutôt de provençal alpin, entrant dans le groupe 19 dits « langues d'oc ». Nous avons également constaté des erreurs (quantitativement marginales) d'affectation des déclarations en Bretagne (où plusieurs dénominations du type *patois Loudéac*, *patois Malesroit* qui relèvent évidemment du « gallo » ont été classées sous « breton ») et en Provence (où des déclarations telles que *rouergue* ou *patois cévenol* ont été classées sous « provençal »).

Les publications qui ont été tirées de l'EHF révèlent à ce sujet une non prise en compte des dénominations et catégorisations des informateurs allant jusqu'à utiliser d'autres dénominations et catégorisations que celles qui ont permis le classement dans la base de données. Ainsi, pour ne prendre qu'un exemple qui nous intéresse ici directement, alors que la classification *Ethnologue*, fiable concernant le domaine d'oc, distingue bien des *langues d'oc* au pluriel (code 19 des données EHF), dont provençal, languedocien, gascon, limousin, auvergnat (divers codes utilisés pour traiter les données EHF), l'ensemble est transformé en un unique *occitan* ou une unique *langue d'oc*, dans Deguillaume et Amrane (2002), dans Héran (2003, 2004), mais pas dans Clanché (2002).

### 2.3. Apports et intérêt de l'enquête linguistique EHF 1999

Au final, cette enquête nous semble pouvoir fournir deux types d'éléments :

- des tendances quantitatives statistiquement valides sur les déclarations de pratiques linguistiques, notamment voire uniquement lorsqu'il s'agit de tendances aux proportions massives et spécifiques ou liées à des facteurs de corrélation significatifs ou inattendus ;
- un certain type de discours épi- et méta-linguistique produit par les informateurs (y compris par les absences de réponses, les écarts de déclarations, etc.), notamment sur la dénomination / catégorisation des langues, c'est-à-dire davantage sur les représentations sociolinguistiques que sur les pratiques linguistiques elles-mêmes.

<sup>1</sup> D'après la liste des codes *Ethnologue* utilisés pour l'EHF, sont également distingués en Suisse le roman de Suisse Romande (scd), le suisse-alémanique (scf), le suisse-italophone qui inclut les Grisons (scj). Or cette liste est erronée et ces codes ne figurent pas dans la base *Ethnologue*, qui classe le franc-comtois dans le français [= incluant les langues d'oïl, code frn], identifie clairement le romanche-grison (code rhe), distinct de l'italien (code itn) et du lombard (code lmo), le suisse-allemand (code gsw) et le romani (code rmo).

S'y ajoute, à un autre niveau :

- un certain type de discours épi- et méta-linguistique produit par les acteurs institutionnels de cette enquête, INED, INSEE, analystes démographes et statisticiens, chercheurs en sociolinguistique, médias... En effet, cette enquête et la publication de certaines de ses exploitations (dont le présent texte) témoignent de représentations sociolinguistiques, scientifiques ou non, et suscitent des discours qui en construisent et révèlent d'autres encore.

Dans tous les cas, les résultats de cette enquête ne nous semblent pas pouvoir être considérés isolément d'autres enquêtes avec lesquelles il convient de les croiser, de les comparer, de les interpréter, afin d'approcher une estimation la plus significative — voire la plus représentative — possible des pratiques linguistiques en France.

### 3. Approches globales des déclarations de pratiques linguistiques en Provence à la fin du XXe siècle

#### 3.1. Approche globale d'un important plurilinguisme

En Provence (cf. 1.2.1. ci-dessus), 20.106 personnes ont répondu à l'enquête EHF et en constituent l'échantillon, sur 4 506 151 habitants (recensement 1999), soit un peu plus de 0,44 % de la population.

##### 3.1.1. Langues autres que français<sup>1</sup> déclarées parfois utilisées pour discuter avec des proches (fig. 4)

Figure 4

Langues (classées par % décroissant)	Proportion de l'échantillon (%)	Extrapolation à l'ensemble de la population (arrondie à l'unité supérieure)
<b>Toutes langues</b>	<b>14,3</b>	<b>644.380</b>
<b>Anglais</b>	<b>4,39</b>	<b>197.820</b>
Italien	2,63	118.512
Espagnol	2,38	107.247
Arabe	2,22	100.037
Provençal	2,2	99.136
Autres langues d'Europe	0,84	37.852
Corse	0,58	26.135
Allemand	0,56	25.235
Autres langues d'Afrique (dont langues berbères)	0,44 (0,24)	19828 10.815
Portugais	0,30	13.519
Langues de Chine, du Tibet, Kmer, Lao, Thaï, Vietnamien	0,23	10.365
Autres langues	0,17	7661
Polonais	0,07	7661
Turc	0,07	7661

On peut remarquer que se détachent principalement 4 langues, si l'on met à part l'anglais dont l'essentiel de la pratique n'est pas dû à une transmission familiale de type régional ou migratoire (cf. fig. 6 et 7) : ce sont l'italien, l'espagnol, l'arabe et le provençal (toutes autour de 2 % de déclarations). Un seuil est marqué par les déclarations d'autres langues, toutes inférieures à 1 %.

##### 3.1.2. Langues autres que français déclarées parfois utilisées pour discuter avec des proches par des habitants de l'agglomération marseillaise (métropole régionale) (fig. 5)

Nous avons pu identifier les informateurs résidants dans l'agglomération marseillaise car parmi les données figure le département et le type d'unité urbaine de résidence, en termes de nombre d'habitants : la métropole régionale est la seule agglomération de cette taille dans les Bouches-du-Rhône.

Les déclarations des informateurs de la métropole régionale confirment la même tendance au plurilinguisme général et portant sur les mêmes langues, avec une part un peu plus

<sup>1</sup> 100 % de l'échantillon est francophone actif.

importante de déclarations de langues portées par des flux migratoires (notamment le corse), et une part un peu moindre du provençal.

**Figure 5**

Langues (classées par % décroissant)	Proportion de l'échantillon (%)	Extrapolation à l'ensemble de la population (arrondie à l'unité supérieure)
<b>toutes langues</b>	<b>20</b>	<b>372.222</b>
<b>anglais</b>	<b>5,60</b>	<b>104.220</b>
arabe	2,96	55.088
italien	2,59	48.202
espagnol	2,40	44.666
provençal	1,64	30.522
corse	1,52	28.288
autres langues d'Europe	0,82	15.260
allemand	0,54	10.050
langues berbères	0,42	7817
autres langues d'Afrique	0,42	7817
autres langues	0,32	5955
portugais	0,29	5397
kmer, lao, thaï, vietnamien	0,28	5211
polonais	0,11	2047
turc	0,06	1117
serbe-croate	0,04	744
langues de Chine et du Tibet	0,02	373

### 3.1.3. Langues autres que français déclarées parlées à un enfant d'environ 5 ans en pourcentages (NB : réponses fournies par l'informateur) par nom de langues (fig. 6)

**Figure 6**

Langues (ordre alphabétique)	Par la mère à l'informateur	Par le père à l'informateur	Par l'informateur à ses enfants
% de répondants	20	20,35	12,55
allemand	0,02	0,04	0
anglais	0,02	0,005	0,11
arabe	3	3	1,23
autres langues	0,2	0,21	0,04
autres langues d'Afrique	0,32	0,3	0,14
autres langues d'Europe	0,96	0,95	0,26
corse	1,08	1,13	0,22
espagnol	2,9	2,71	0,81
italien (+ dialectes)	4,6	4,11	0,79
kmer, lao, thaï, vietnamien	0,3	0,25	0,1
langues berbères	0,35	0,35	0,13
langues de Chine et du Tibet	0,01	0,01	0,01
polonais	0,06	0,2	0,04
portugais	0,4	0,4	0,19
provençal	4,62	5,87	0,98
langues régionales	25	42	17
serbe-croate	0,08	0,06	0,02
turc	0,1	0,1	0,05

### 3.1.4. Langues autres que français déclarées parlées à un enfant d'environ 5 ans en pourcentages (NB : réponses fournies par l'informateur) par ordre décroissant de déclaration (fig. 7)

Le provençal est la langue dont l'usage parents > enfants déclaré par les enfants devenus adultes a été le plus élevé au cours du XXe siècle par rapport à l'ensemble des langues autres que le français. Suivent l'italien et autres variétés linguistiques italiennes (conséquence d'un

forte immigration italienne en Provence entre 1850 et 1950), l'arabe, l'espagnol et, moins représenté, le corse. On tombe ensuite à des taux de déclaration beaucoup plus bas (inférieurs à 1 %).

**Figure 7**

Langues	Parlée à l'informateur par père et/ou mère en %
<b>% de répondants</b>	<b>20,175</b>
provençal	5,245
italien (+ dialectes)	4,355
arabe	3
espagnol	2,805
corse	1,105
autres langues d'Europe	0,955
portugais	0,4
langues berbères	0,35
autres langues d'Afrique	0,31
kmer, lao, thaï, vietnamien	0,275
autres langues	0,205
polonais	0,2
turc	0,1
serbe-croate	0,08
allemand	0,03
anglais	0,02
langues de Chine et du Tibet	0,01

### 3.1.5. Langues déclarées parlées par l'informateur à ses enfants âgés d'environ 5 ans par ordre décroissant de déclarations (fig. 8)

**Figure 8**

Langues	Parlée par l'informateur à ses enfants en %
<b>% de répondants</b>	<b>12,55</b>
arabe	1,23
provençal	0,98
espagnol	0,81
italien (+ dialectes)	0,79
autres langues d'Europe	0,26
corse	0,22
portugais	0,19
autres langues d'Afrique	0,14
langues berbères	0,13
anglais	0,11
kmer, lao, thaï, vietnamien	0,1
turc	0,05
autres langues	0,04
polonais	0,04
serbe-croate	0,02
langues de Chine et du Tibet	0,01
allemand	0

Les usages parents > enfants déclarés par les parents atteignent des proportions beaucoup plus basses : moitié moins de déclarations au total. Mais ce sont, à l'exception du corse, les mêmes langues qui viennent en tête des déclarations, dans un ordre modifié. Trois facteurs peuvent en rendre compte :

- il est plus acceptable d'attribuer des pratiques linguistiques réputées « anciennes », voire dévalorisées, à ses parents qu'à soi-même,

- les générations les plus récentes déclarent utiliser beaucoup moins les langues en question avec leurs propres enfants,
- le mouvement migratoire arabophone est le plus récent parmi ceux concernés ici, et la migration hispanophone est plus récente que la migration italophone (d'où un décalage des générations).

On note de ce point de vue une place des déclarations relativement marquée pour le provençal, qui se maintiennent au 2e rang, mais dont l'écart avec les déclarations concernant les parents des informateurs est également le plus marqué : indices d'une présence sociolinguistique certaine accompagnée d'une certaine disqualification sociale et d'une forte érosion des usages parents > enfants.

### **3.2. Les axes statistiquement discriminants des déclarations de plurilinguisme en Provence**

#### **3.2.1. Langues étrangères et de l'immigration en Provence**

En ce qui concerne les langues étrangères et de l'immigration, l'étude des valeurs propres permet de déterminer le nombre adéquat d'axes à étudier. Le diagramme des proportions d'inertie expliquée par les axes sur le tableau de Burt, confirmé par les taux de Greenacre montre que, au-delà du troisième axe, les axes peuvent être considérés comme porteurs de bruit statistique (leur taux de Greenacre associé ne dépasse pas les 5 %). Les trois premiers axes ont donc été retenus et nous présentons ici les résultats de l'axe 1, le plus significatif, qui représente 9,84% de l'inertie totale et dont on reconstitue 82,23 % de l'information avec les modalités retenues. Il suggère une opposition statistique entre :

- a) des jeunes de nationalité française dont la langue maternelle unique est le français,
- b) des personnes âgées ayant reçu presque exclusivement de leurs parents une langue de l'immigration (espagnol, italien, arabe).

D'un côté, on a affaire à des personnes dont les parents sont nés en France tout comme elles. Ce sont des personnes qui ont des diplômes de niveau supérieur à bac+2, elles sont jeunes (moins de 30 ans) et n'ont pas d'enfants. Ces individus sont issus de familles de taille moyenne (deux frères et sœurs) et vivent dans la région de leur enfance. Concernant leur travail, ils sont tout comme leurs parents soit dans la fonction publique (santé, enseignement, professions intellectuelles, artistiques...) soit cadres d'entreprises. Quant aux langues, ils ont en majorité le français pour unique langue maternelle, bien que le provençal tienne une place importante en tant que langue occasionnelle reçue des parents. En revanche, cette dernière langue n'apparaît plus lorsque l'on s'intéresse aux langues transmises par l'informateur à ses enfants et à celles parlées avec les proches.

De l'autre côté de l'axe, ce sont des individus de plus de 50 ans, sans aucun diplôme, tout comme leur conjoint(e), ouvriers ou employés à la retraite (alors que leur conjoint(e) est « au foyer ») dont le père était ouvrier agricole. Ils se distinguent également par le nombre d'enfants qu'ils ont eu (plus de quatre) et le fait qu'ils sont eux aussi issus de familles nombreuses (plus de sept frères et sœurs). Ce sont des Français par acquisition ou des étrangers dont les parents sont nés soit en Italie, soit en Espagne et qui ont déclaré avoir reçu presque exclusivement d'eux l'italien (ou un « dialecte » italien) ou l'espagnol. En revanche, même si ces deux langues sont bien représentées lorsqu'il s'agit des langues transmises par l'informateur à ses enfants, il faut noter qu'elles le sont bien moins lorsqu'on s'intéresse aux langues parlées avec les proches. Il faut aussi remarquer la présence constante de l'arabe et des langues berbères.

Cet axe confirme la première lecture des données comme quoi les langues de l'immigration les plus fréquemment parlées en Provence sont l'italien, l'espagnol et l'arabe. En revanche, les deux premières langues sont bien mieux représentées que l'arabe à gauche de l'axe, car à des jeunes français de naissance ayant reçu de leurs parents presque exclusivement le français, on oppose des langues de l'immigration reçues par des gens plus âgés. Ainsi, cet axe permet de mettre en avant l'idée que les personnes ayant reçu l'italien ou l'espagnol de leurs parents sont plus âgées que la moyenne régionale (il ne faut toutefois pas confondre langues reçues et langues parlées, où l'italien et l'espagnol sont proches de la moyenne régionale). L'arabe résiste mieux aux déclarations de passage au monolinguisme français. En effet, moins de 40 %

des personnes ayant reçu d'au moins l'un de leurs parents l'arabe et ayant des enfants ont déclaré ne leur avoir transmis que le français, alors que pour l'espagnol et l'italien ont atteint respectivement 62 % et 74 %. Mais cela est sans doute aussi dû au fait que l'immigration maghrébine en France est plus récente que l'espagnol ou l'italienne. L'information principale de cet axe est que les jeunes Français de naissance dont les parents sont eux-mêmes nés en France ont moins tendance à être en contact avec une langue autre que le français, à l'exception partielle du provençal, qui apparaît de ce côté-là car ceux qui le déclarent sont surtout des personnes de nationalité françaises dont les deux parents sont nés en France : 95 % des personnes déclarées comme parlant le provençal sont de nationalité française et parmi eux, 90 % ont des parents nés en France. Mais cela signifie également que, parmi les personnes déclarées comme parlant provençal, 5 % n'ont pas la nationalité française et 10 % ont des parents nés à l'étranger.

Les axes 2 (4,84 % de l'inertie totale et 79,57 % de l'information) et 3 (3,99 % de l'inertie totale et 81,55 % de l'information) suggèrent, en résumé :

- Axe 2 : Une opposition statistique entre transmission déclarée de plus en plus faible de l'italien (ou d'un « dialecte » italien), de génération en génération, chez des Italiens qui ont acquis la nationalité française et une pratique du français déclarée en baisse chez les jeunes immigrés sans diplôme provenant du Maghreb avec un taux de transmission déclarée de l'arabe ou du berbère en hausse par rapport aux générations plus anciennes. Si 77 % des personnes de plus de 60 ans dont les parents sont nés au Maghreb déclarent n'avoir reçu que le français (ou l'arabe occasionnellement), ils ne sont plus que 57 % chez les moins de 30 ans. Dans le même temps, 19 % des plus de 60 ans déclarent avoir reçu vers 5 ans l'arabe de façon exclusive ou presque, alors que ce taux atteint 33 % chez les moins de 30 ans (dont une bonne partie est née dans un Maghreb en phase d'arabisation).
- Axe 3 : Une opposition statistique entre des personnes âgées déclarant parler italien (ou un « dialecte » italien) et des personnes d'origines maghrébines, plus jeunes, dont la transmission déclarée de l'arabe (ou du berbère) à leurs enfants est seulement occasionnelle. D'un côté, nous avons des individus âgés (plus de 60 ans), mariés, sans enfants, et dont le ou la conjoint(e), tout comme leurs parents, sont nés en Italie. Ce sont des retraités (anciens ouvriers ou employés, anciens agriculteurs ou encore anciens artisans ou commerçants), de niveau certificat d'études primaires. De l'autre côté, ce sont des personnes dont le (ou la) conjoint(e) et les parents sont nés au Maghreb, couples de quadragénaires avec deux ou trois enfants, et où seul l'homme travaille, soit comme ouvrier qualifié, soit comme ouvrier agricole. Tous deux ont le CAP ou le BEP. Ils sont issus de familles nombreuses, n'ont pas la nationalité française et vivent de préférence dans des villes moyennes (de 20000 à 50000 habitants) des Bouches-du-Rhône.

### 3.2.2. Les langues régionales en Provence

L'analyse de l'histogramme des valeurs propres et celle de la représentation des modalités ayant un rapport avec les langues régionales amènent à ne retenir que 3 axes (analyses factorielles des correspondances multiples sans variables supplémentaires). Les variables concernant les langues régionales ont été portées en supplémentaires : on note l'importance d'un l'axe 1 qui, d'après le taux de Greenacre, restitue à lui seul 25 % de l'information, puis un saut relatif après l'axe 3. Il convient donc de limiter l'étude à 3 axes ( $> 0,05$ ). L'AFCM sans variables supplémentaires permet de faire ressortir trois facteurs déterminants.

Selon le premier facteur, les modalités concernant le provençal alpin sont bien représentées sur l'axe et sont les plus contributives, avec celles concernant le département de résidence des Hautes-Alpes (CTR = 1,89 %) et département de naissance (CTR = 2,37%). Les modalités « provençal déclaré parlé par le père » et « provençal déclaré parlé par la mère » sont elles aussi bien représentées mais moins contributives. Les modalités les plus contributives restent celles concernant le niveau d'étude et l'origine sociale des parents et des individus.

Le deuxième facteur est caractérisé par les locuteurs déclarés du corse, individus plutôt âgés, retraités, vivant dans des zones urbaines des Bouches du Rhône (notamment à Marseille). Ils déclarent avoir reçu le corse de leurs deux parents et le parler avec des proches mais qui ne pas l'avoir parlé à leurs enfants quand ceux-ci avaient 5 ans. Ils sont statistiquement opposés à des locuteurs du provençal alpin.

Les regroupements sur le troisième facteur s'opèrent selon des critères géographiques puis sociaux. La principale information concerne des individus déclarant avoir reçu le provençal de leurs parents, le parler avec des proches et l'avoir transmis à leurs enfants. Il s'agit d'ouvriers ou d'employés vivant dans les Alpes de Haute Provence, dans le Var ou dans le Vaucluse, âgés de 50 à 70 ans et habitant des zones urbaines de taille moyenne.

L'axe 1 (25 % de l'information), ainsi déterminé, oppose statistiquement les agriculteurs aux professions intermédiaires et aux cadres.

D'un côté on a : des anciens agriculteurs avec conjoint agriculteur aujourd'hui sans enfant, tous nés et résidant dans les Hautes-Alpes, âgées de 70 à 95 ans, ayant un niveau d'étude faible, issus de parents agriculteurs. Ce sont les déclarants privilégiés de la pratique du provençal alpin. Le provençal alpin apparaît quasiment anecdotique hors des Hautes-Alpes<sup>1</sup>. Il s'agit d'individus qui n'ont pas reçu le français de leurs parents ou de façon occasionnelle et qui ont transmis le provençal alpin de façon habituelle à leurs propres enfants.

De l'autre on a : des familles monoparentales et des couples sans enfants ou avec un à deux enfants, âgés de moins de 40 ans, de niveau d'étude élevé, occupant des professions intermédiaires ou de cadres, issus de parents présentant les mêmes caractéristiques. On y trouve les déclarations d'usage de langues étrangères comme l'anglais, majoritairement utilisées par les cadres dans un contexte professionnel. Les locuteurs déclarés de langues de l'immigration, marginaux dans cette population, n'apparaissent pas sur cet axe. Les locuteurs déclarés d'autres langues régionales sont aussi bien représentés de ce côté de l'axe. Il s'agit en effet d'une population plutôt urbaine et originaire d'autres régions. On y trouve notamment les modalités « corse déclaré parlé par le père » et « corse déclaré parlé par la mère » mais les modalités « corse déclaré parlé avec des proches » « et corse déclaré parlé aux enfants à 5 ans » ne sont pas significatives. Il s'agit d'individus ayant reçu le corse de façon occasionnelle et qui ne pratiquent donc plus la langue. Ce sont pour la plupart des cadres ou de professions intermédiaires, de parents cadres ou profession intermédiaire.

L'axe 2 (8,47 % de l'information), met en évidence le fait qu'on trouve une forte proportion d'employés et d'ouvriers, habitant plutôt dans les régions rurales parmi les informateurs qui déclarent employer le provençal. Ils déclarent avoir reçu le provençal (y compris le provençal alpin) de façon habituelle. Les modalités « langue déclarée parlée par le père », « langue déclarée utilisée avec des proches » et « langue déclarée parlée aux enfants de 5 ans » concernant le provençal sont en effet significatives. Ce sont des individus âgés de 30 à 54 ans, de niveau CAP/BEP, vivant en couple avec 1 ou 2 enfants.

À l'opposé on trouve des individus pratiquant une autre langue régionale ou une langue étrangère (les cadres, dans le milieu professionnel, cf. axe 1). Il s'agit d'individus jeunes (18-24ans) ou à la retraite (65 à 79 ans), vivant dans des grandes villes, qui déclarent avoir reçu une autre langue que le français de façon occasionnelle. La discrimination sur cet axe se fait avant tout par le niveau d'étude et le groupe social (la modalité « cadre » a une CTR de 5,27 %, la modalité « niveau d'étude 4 » une CTR de 6,28 %).

Ainsi, les individus habitant dans des communes rurales et déclarant plus utiliser le provençal que les autres, sont principalement ouvriers, voire employés. Ils déclarent avoir reçu la langue régionale de façon habituelle, davantage par leurs pères, l'avoir parlée à leurs enfants et l'employer avec des proches.

Enfin, l'axe n° 3 (7,62 % de l'information) caractérise des individus nés et vivant dans le Var et les Bouches du Rhône, dans les métropoles urbaines (Marseille, Toulon). Il s'agit d'individus ouvriers, ayant un niveau d'étude « faible », dont le père et/ou la mère sont ou étaient ouvriers, ou la mère est ou était au foyer. Les modalités « corse reçu de père/mère », « corse transmis aux enfants » et « corse parlé avec des proches » sont significatives de ce côté de l'axe, tout comme les modalités correspondant aux autres langues régionales et étrangères (cf. axe 1 ci-dessus). Il s'agit ici d'individus qui déclarent que leurs parents leur ont parlé à 5 ans une autre langue que le français de façon habituelle, voire que leurs parents ne leur ont pas du tout transmis le français (surtout leur mère), et qui déclarent n'avoir eux-mêmes transmis à leurs enfants qu'une seule autre langue et pas le français ou alors de façon seulement occasionnelle.

---

<sup>1</sup> Mais c'est là peut-être aussi un effet du regroupement linguistique par recodage opéré par l'INSEE.



À l’opposé, l’axe caractérise des hommes vivant dans le Vaucluse, et dans des communes rurales de Provence. On retrouve à la fois des cadres, des professions intermédiaires et des agriculteurs (effet Guttman). Les modalités concernant le monde agricole y sont fortement contributives (père agriculteur : CTR = 5 %, mère agricultrice : CTR = 5,59 %, anciens agriculteurs : CTR = 5,50 %, agriculteurs : CTR = 7,65 %) et sont bien représentées. Il s’agit d’individus déclarant avoir reçu le provençal de leurs parents de façon habituelle et avoir eux-mêmes parlé le provençal à leurs enfants.

En conclusion, nous pouvons constater que le provençal alpin est une langue régionale dont les pratiques déclarées s’avèrent fortement corrélées à des facteurs sociaux et géographiques locaux. Les pratiques déclarées du provençal dans le reste de la Provence le sont beaucoup moins. Une forte proportion d’agriculteurs l’utilise et cela reste une langue rurale mais ces caractéristiques sont moins marquées que dans d’autres régions que nous avons étudiées par ailleurs (Bretagne historique et Normandie historique, cf. *supra*). Il n’en reste pas moins qu’il y a des zones où l’on déclare davantage de pratiques du provençal que d’autres. Une plus forte proportion d’hommes que de femmes déclare en faire usage, même si cela est beaucoup plus marquant pour le provençal alpin que pour le provençal. Les déclarations suggèrent un recul des pratiques et notamment de la transmission familiale du provençal dans toute la région considérée, pratiques qui concerneraient surtout des individus assez âgés.

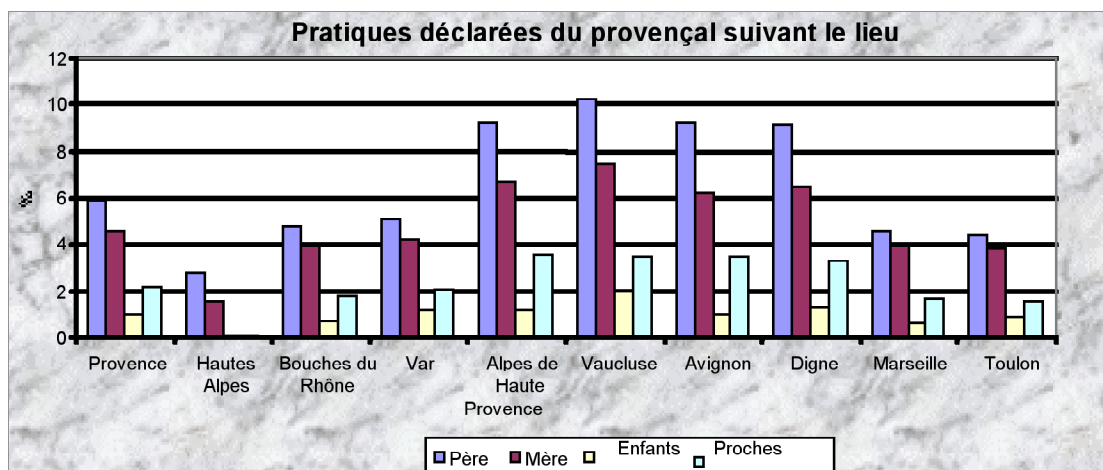
### 3.3. Informations complémentaires et synthétiques sur les pratiques et représentations du provençal

Des analyses précédentes ressortent déjà diverses tendances concernant le provençal, langue évidemment la plus caractéristique de la situation sociolinguistique de la Provence, non seulement par sa spécificité locale (toutes les autres langues attestées en Provence le sont également ailleurs et souvent de façon plus importante alors que le provençal n’est fortement attesté qu’en Provence), mais aussi par la place qu’il y occupe : il tient le premier rang des langues déclarées parlées par les parents des informateurs et se maintient dans le groupe des quatre langues les plus déclarées à côté du français et de l’anglais dans la région. Un certain nombre de croisements permettent de les compléter et de les préciser, sans pour autant entrer ici dans une exploitation sociolinguistique très détaillée de l’EHF.

#### 3.3.1. Pratiques déclarées du provençal (fig. 9 et 10)

NB : Les éléments de légende père, mère, enfants, proches renvoient respectivement aux questions précises n° 19 et 20 de l’EHF

Figure 9



Tous lieux<sup>1</sup> et toutes générations confondues, selon les informateurs, c’est donc 13,65 % de la population qui sont déclarés concernés par les types de pratiques limitées de provençal recherchées par l’enquête. Les 2,2 % d’informateurs déclarant discuter parfois en provençal avec proches représenteraient ainsi 99.136 personnes. Cette estimation d’environ 100.000 locuteurs actuels du provençal en Provence au sens ici retenu (il faudrait y ajouter ceux des Alpes-Maritimes, du Gard, de la Drôme méridionale et résidant ailleurs), est inférieure à tous

<sup>1</sup> Pour la détermination des villes de résidence, nous avons appliqué la même méthode que pour identifier Marseille (cf. *supra*).

les résultats d'enquêtes récentes. Selon une enquête médiamétrie commandée par France 3 fin 1998, d'un type directement comparable à l'EHF, dans deux zones très urbanisées et à la population très diversifiée (donc peu propices à la langue locale), il y aurait dans l'agglomération marseillaise 23 % de locuteurs passifs (300.000 personnes) et 5 % d'actifs (65.000), sur la côte varoise 17 % de passifs (70.000 personnes) et 4 % d'actifs (16.000), dont 70 % du tout ont plus de 50 ans et environ 10 % moins de 25 ans (Blanchet, 2002, pp. 20 et suiv.). On voit qu'on est au-dessus des 1,6 % à 2 % de pratiques actives déclarées dans les mêmes zones lors de l'EHF.

**Figure 10**

Déclaré parlé...	Par le père	Par la mère	aux enfants	avec les proches	% total de personnes concernées
Provence	5,87	4,62	0,98	2,2	13,67
Hautes Alpes	2,72	1,54	0,12	0,12	4,5
Alpes de Haute Provence	9,21	6,71	1,17	3,51	20,6
Bouches du Rhône	4,84	3,98	0,67	1,79	11,28
Var	5,16	4,27	1,21	2,08	12,72
Vaucluse	10,19	7,47	1,99	3,4	23,05
Avignon	9,23	6,23	1	3,49	19,95
Digne	9,19	6,46	1,24	3,23	20,12
Marseille	4,63	3,89	0,58	1,64	10,74
Toulon	4,46	3,84	0,88	1,6	10,78

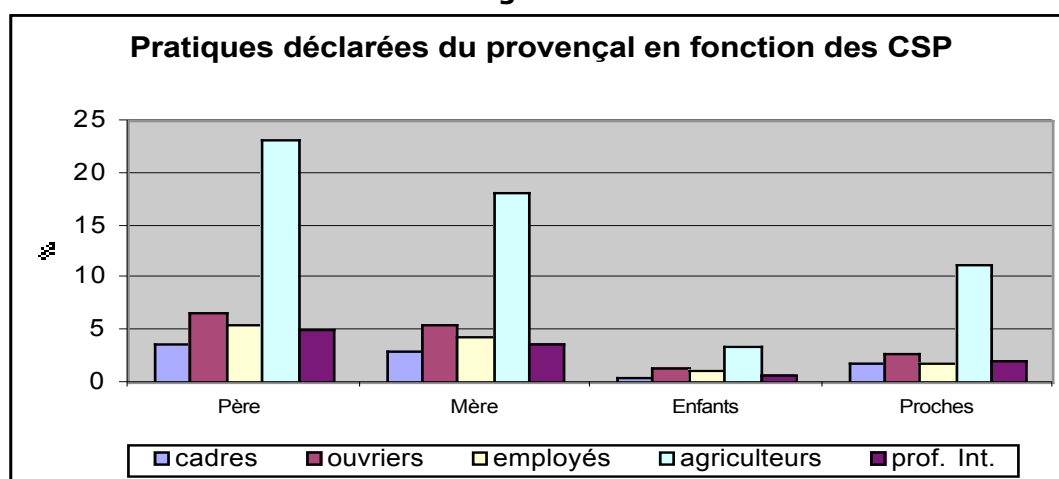
Toutes formes de pratiques déclarées confondues, les 100.000 usagers représentés à qui « il arrive de discuter en provençal avec leurs proches » sont probablement inclus dans les 250.000 locuteurs réguliers estimés par croisement de toutes les sources d'information disponibles (Blanchet, *ibid.*) et qui utiliseraient le provençal dans des circonstances plus larges ou qui le déclareraient dans des circonstances plus favorables :

**Figure 11 : Estimations démolinguistiques croisées concernant le provençal**

Habitant la région administrative en 1999	4, 5 millions
dont 80 % se déclarant attachés au provençal	3,6 millions
et dont la moitié installés depuis + de 35 ans <sup>1</sup>	2, 25 millions
dont la moitié de locuteurs (passifs compris)	1 million
dont la moitié de locuteurs actifs (occasionnels compris)	500 000
dont la moitié de locuteurs réguliers	250 000

### 3.3.2. Répartition par CSP (fig. 12)

**Figure 12**



<sup>1</sup>Dont seulement 400 000 étrangers, dont seulement 250 000 hors CEE. Il y a donc plus d'1,5 millions de *Franchimand* "Français non méridionaux", diffuseurs directs et massifs du français dominant. On peut estimer que seulement 2 millions d'habitants ont des parents et grands-parents nés en Provence mais qu'environ 250 000 Provençaux vivent en diaspora, auxquels il faut ajouter les provençalophones de la Drôme, du Gard, voire d'Italie, non comptabilisés ici. Source INSEE.

La pratique du provençal est avant tout déclarée par / pour les agriculteurs. Mais toutes les Catégories Socio-Professionnelles sont concernées de façon relativement comparable entre elles, même si on observe une baisse à mesure que le « niveau social » augmente. Les analyses factorielles présentées ci-dessus indiquent notamment que les catégories urbaines des ouvriers et employés en sont également déclarées usagers principaux.

### 3.3.3. Noms donnés aux pratiques linguistiques régionales de Provence (≈ au « provençal ») par les informateurs déclarant l'employer avec leurs proches (fig. 13)

Figure 13

Dénominations	Pourcentage d'informateurs
patois	1,9
patois provençal	8,1
provençal	84,3
provence	0,5
langue d'oc	0,26
occitan	5

Les chiffres donnés ici sont significatifs uniquement pour les tendances lourdes, car les éventuels problèmes de recodage demandent, notamment pour les dénominations ambiguës (*patois*, *langue d'oc*, *occitan*, 22 occurrences) des vérifications informateur par informateur. Ainsi, sur les 11 déclarant en première position l'*occitan* comme autre langue que le français employé avec des proches (et recodés par l'INED comme usagers du provençal...), 1 seul a des parents nés en Provence, 6 ont des parents (père, mère ou les deux) nés dans d'autres régions (Ardèche, divers départements du sud-ouest), 3 ont des parents nés dans les Hautes-Alpes (où l'identité linguistique est moins nettement provençale, on l'a vu) et 1 les déclare nés « en France » sans plus de précision. Un enquêté a des parents nés à l'étranger (probablement en Italie, et donc dans une zone linguistique italienne de langue d'oc, puisqu'il déclare aussi l'italien comme langue familiale). Plus de la moitié de ces déclarations ne recouvre donc pas le provençal.

11 autres déclarations ne sont pas comptabilisées ici, puisque ces 11 informateurs, résidant en Provence, ont indiqué clairement *patois gascon*, *patois languedocien*, *patois Midi Pyrénées*, etc., — encore recodés comme « provençal » par l'INED...<sup>1</sup>, c'est-à-dire des dénominations qui renvoient clairement à autre chose que du provençal.

Le terme *Provence* est probablement une spécification des termes *langue*, *dialecte* ou *patois* proposés par le questionnaire. Trois tendances apparaissent néanmoins clairement.

- a- La dénomination spontanée déclarée par les informateurs est donc massivement *provençal*, tendance lourde observée depuis longtemps, bien qu'il s'agisse là d'un point qui fasse l'objet de contestations (notamment par une certaine militance en faveur d'un *occitan* globalisant) et en retour, à l'inverse, de soutiens très vifs dans le corps social, les milieux associatifs, les collectivités territoriales et les élus, puisque le statut sociolinguistique du provençal et les mesures de politique linguistique qui en découlent sont directement concernées (cf. Blanchet, 2004). Si l'on y ajoute les spécifications *patois provençal* / *de Provence*, on atteint les 92 % de dénominations qui rattachent ainsi la langue à l'espace régional.
- b- Les dénominations à tendance éventuellement péjorative, comme *patois* (10 %) ou *dialecte* (absent), sont très peu usitées dans ces déclarations.
- c- Les dénominations englobantes (*langue d'oc*, *occitan*) sont encore moins usitées (environ 5 %), sachant que la majorité d'entre elles réfèrent en fait à des parlers d'autres régions dont seraient originaires les informateurs.

<sup>1</sup> Fonctionnellement, cela montre que le regroupement, à un certain niveau, des variétés linguistiques gallo-romanes du sud de la France en un seul *occitan*, et même en un seul groupe de *langues d'oc*, induit des erreurs d'analyse.

#### 4. Quelques points de comparaison pour ne pas conclure...

Ce travail d'exploitation de l'EHF a permis de soulever une série de problèmes méthodologiques qui relativisent les résultats de cette enquête et suggèrent d'autres enquêtes selon des protocoles plus adaptés (le prochain recensement pourrait opportunément en offrir la possibilité). Globalement, d'ailleurs, l'EHF ne fait que confirmer statistiquement des tendances déjà observées sur le terrain par des enquêtes diverses à plus petites échelles, sauf exception.

La situation sociolinguistique de la Provence, fortement plurilingue telle qu'elle est représentée selon les déclarations des informateurs, est marquée par quatre ensembles de pratiques linguistiques :

- des pratiques générales du français (dont les variations locales et régionales en Provence ont été étudiées, cf. notamment Binisti et Gasquet-Cyrus, 2001, Blanchet 2002) ;
- des pratiques professionnelles de l'anglais qui sont tout aussi minoritaires que celles des autres langues (à peine plus de 5 % de déclarations d'usage, moins de 0,1 % de déclarations de transmission) ;
- des pratiques apparemment minoritaires et néanmoins significatives du provençal d'une part et de langues issues de migrations marquantes d'autre part (italien, espagnol, arabe, et corse à Marseille) ;
- et enfin des pratiques très minoritaires de nombreuses autres langues apportées par divers flux migratoires d'origines internes ou externes à la France.

La généralisation du français y apparaît comme un fait acquis. L'appropriation de cette langue sous une forme provençalisée, largement partagée comme marqueur identitaire, en est un indicateur qualitatif complémentaire : le français régional joue désormais un rôle identitaire comparable à celui que jouent des langues régionales, à d'autres époques ou dans d'autres régions, lorsqu'elles sont à la fois massivement parlées et variétés basses d'une diglossie (Binisti et Gasquet-Cyrus, 2001). L'évolution vers une diminution du plurilinguisme au profit du seul français dominant, tant chez les locuteurs déclarés ou potentiels du provençal que chez ceux de langues issues de l'immigration, est un phénomène largement mis en avant par les informateurs, notamment au regard de croisements avec le facteur « âge ». L'italien y a joué et semble continuer malgré tout à y jouer un rôle notable, davantage que dans la plupart des régions françaises (d'autant que les Alpes-Maritimes ont ici été soustraites de nos analyses).

Les principales corrélations sociologiques des usages déclarés de ces langues sont mises en relief par les analyses factorielles décrites ci-dessus. Il est notamment frappant de faire la constatation, plutôt contre-intuitive, que le provençal est probablement la seule langue autre que le français qui concerne l'ensemble des secteurs de la population de la région, toutes nationalités, toutes générations, toutes catégories socio-professionnelles et tous lieux confondus, même si, bien sûr, certains sous groupes de l'échantillon en sont déclarés davantage usagers que d'autres : les agriculteurs, ouvriers et employés des zones rurales et semi-urbaines hors des Bouches-du-Rhône. C'est un facteur qui contribue probablement à son investissement symbolique massif comme élément identitaire du paysage sociolinguistique régional.

Si l'on compare les résultats de l'EHF pour la Provence aux résultats nationaux (hors DOM-TOM), on constate que son taux de plurilinguisme ici déclaré se situe légèrement sous la moyenne, avec 20 % de déclarations d'usages d'autres langues que le français par les parents des informateurs et par les informateurs eux-mêmes, contre 26 % pour les parents et 21 % pour les informateurs au niveau national. Ce taux d'informateurs usagers est plus élevé encore en Picardie (près de 30 %, notamment grâce à l'immigration) mais moins élevé en Normandie (9,75 %) et, en ce qui concerne les langues de l'immigration, en Bretagne historique (3 %). Proportionnellement, le provençal est beaucoup moins déclaré (2,2 % d'informateurs usagers) que le corse (45 %), l'alsacien (39 %), le breton (15 %), ou le basque. Même chose par rapport aux variétés d'oc, puisqu'on note par exemple 7 % d'informateurs usagers déclarés de l'occitan en Aquitaine (les chiffres ne sont pas encore connus pour les autres régions du domaine linguistique d'oc). Mais, à côté d'une extrapolation à hauteur d'à peine 526.000 locuteurs supposés sur l'ensemble des pays d'oc (une trentaine de départements sur en gros 6 régions administratives), la Provence fournit à elle seule un cinquième des effectifs (100.000 locuteurs actifs ici supposés) et maintient une présence significative. Inversement, le provençal est davantage déclaré que le gallo (environ 1,5 % d'informateurs se déclarant usagers) ou

que le normand (1 %). Il faut évidemment rapporter ces chiffres à des situations démologiques différentes : la Provence est une région ayant connu depuis les années 1960 une véritable explosion démographique (les locuteurs potentiels, « héritiers » du provençal, n'y représentent plus que moins de la moitié de la population), ce qui n'est pas le cas de la Normandie, ni même de la Bretagne.

On notera aussi la très faible place du portugais (2e langue de l'immigration déclarée en France métropolitaine après l'arabe), du polonais et du turc, beaucoup mieux représentés au niveau national.

Et d'autres analyses de ce volet linguistique de l'EHF 1999 restent en attente...

## Références bibliographiques

- Binisti (N.) & Gasquet-Cyrus (M.). 2001. *Le français de Marseille, description sociolinguistique*. Aix-en-Provence : DGLF, Université de Provence [Rapport de recherche].
- Blanchet (P.). 1995. « Note de lecture : Fañch Broudic, *La Pratique du breton de l'ancien régime à nos jours* ». in : *La Linguistique*, 31, 1995-2, pp. 153-158. Rennes : Presses Universitaires de Rennes.
- Blanchet (P.). 2000. *Linguistique de terrain, méthode et théorie (une approche ethno-sociolinguistique)*. Rennes : Presses Universitaires de Rennes.
- Blanchet (P.). 2002. *Langues, cultures et identités régionales en Provence*. Paris : L'Harmattan.
- Blanchet (P.). 2003a. « Contacts et dynamique des identités culturelles : les migrants italiens en Provence dans la première partie du XXe siècle ». in : *La France Latine - revue d'études d'oc*, 137, pp. 141-166. Paris, 2003 [Article mis en ligne dans le n° 6 de la revue électronique *La-Science-Politique.com*].
- Blanchet (P.). 2003b. « Compte rendu de F. Héran, A. Filhon et Ch. Deprez, *La dynamique des langues en France au fil du XXe siècle - Population & Société* (bulletin de l'INED, octobre 2001) et François Clanché, *Langues régionales, langues étrangères : de l'héritage à la pratique - Bulletin de l'INSEE n° 830, février 2002*. ». in : *Marges Linguistiques*, 5, pp. 22-24 [<http://www.marges-linguistiques.com>].
- Blanchet (P.). 2004. « Provençal as a distinct language ? : sociolinguistic patterns revealed by a recent public and political debate ». in : Blanchet (P.) & Schiffman (H.), (dir.). 2004. « The Sociolinguistics of Southern « Occitan » France Revisited ». in : *International Journal of the Sociology of Language*, 169, pp. 125-150. Berlin, New-York : Mouton de Gruyter.
- Blanchet (P.) & Schiffman (H.), (dir.). 2004. « The Sociolinguistics of Southern « Occitan » France Revisited ». in : *International Journal of the Sociology of Language*, 169, pp. 125-150. Berlin, New-York : Mouton de Gruyter.
- Blot (D.), Eloy (J.-M.) & Rouault (T.). 2004. « La richesse linguistique du nord de la France ». in : *Relais*, 125. Amiens : INSEE-Picardie.
- Broudic (F.). 1995. *La pratique du breton de l'ancien régime à nos jours*. Rennes : Presses Universitaires de Rennes.
- Bulot (T.), Bauvois (C.) & Blanchet (P.), (dir.). 2001. *Sociolinguistique urbaine. Variations linguistiques : images urbaines et sociales*. Rennes : Presses Universitaires de Rennes.
- Bulot (T.) & Messaoudi (L.), (dir.). 2003. *Sociolinguistique urbaine : frontières et territoires*. Cortil-Wodon : Éditions Modulaires Européennes.
- Calvet (L.-J.). 1994. *Les voix de la ville*. Paris : Payot.
- Calvet (L.-J.). 2004. *Essais de linguistique. La langue est-elle une invention des linguistes ?* Paris : Plon.
- Cassan (F.), Héran (F.) & Toulemon (L.). 2000. « Étude de l'histoire familiale, l'édition 1999 de l'enquête famille ». in : *Courrier des Statistiques*, 93, pp. 25-37.
- Cesari (J.), Moreau (A.) & Schleyer (A.). 2001. *Plus marseillais que moi, tu meurs ! Migrations, identités et territoires à Marseille*. Paris : L'Harmattan.
- Clanché (F.). 2002. « Langues régionales, langues étrangères : de l'héritage à la pratique ». in : *Bulletin de l'INSEE*, 830. Paris : INSEE.
- Dalbera (J.-P.). 1994. *Les parlers des Alpes-Maritime*. Londres : AIEO.
- Deguillaume (C.) & Amrane (E.). 2002. « Langues parlées en Aquitaine : la pratique héritée ». in : *Le Quatre Pages*, 110. Bordeaux : INSEE-Aquitaine.
- Dubarry (B.) & Dupouts (D.), (ed.). 1995. *Pratique, présence et représentations de l'Occitan dans les Hautes-Pyrénées. Enquête sociolinguistique*. Tarbes : Conseil Général des Hautes-Pyrénées.
- Duée (M.). 2002. « L'Alsacien, deuxième langue régionale de France ». in : *Chiffres pour l'Alsace*, 12. Strasbourg : INSEE-Alsace.
- Eloy (J.-M.), (ed.). 1998. *Évaluer la vitalité des variétés d'oïl et autres langues*. Amiens : Centre d'Études Picardes de l'Université d'Amiens.
- Hammel (E.) & Gardy (P.). 1994. *L'occitan en Languedoc-Roussillon. 1991*. Perpignan : El Trabucaire.
- Héran (F.). 1993. « L'unification linguistique de la France ». in : *Populations & Société*, 285. Paris : INED.

- Héran (F.). 2003. « La situation des langues régionales en France ». in : *Actes des 1ères Assises nationales des langues de France*, pp. 5-6. Paris : Ministère de la Culture.
- Héran (F.). 2004. « Une approche quantitative de l'intégration linguistique en France ». in : *Langues de France : Hommes et migrations*, 1252, pp. 10-24.
- Héran (F.), Filhon (A.) & Deprez (C.). 2001. *La dynamique des langues en France au fil du XXe siècle - Population & Société*. Paris : INED.
- Laurent (L.). 1992. « Komz a rit a brezhoneg ? (parlez-vous breton ?) ». in : *Octant*, 48. Rennes : INSEE-Bretagne.
- Le Boëtté (I.). 2003. « Langue bretonne et autres langues : pratique et transmission » in : *Octant*, 92, pp. 18-22. Rennes : INSEE-Bretagne.
- Marien (B.) & Beaud (J.-P.). 2003. *Guide pratique de l'utilisation de la statistique en recherche : le cas des petits échantillons*. Québec : AUF.
- Oyharçabal (B.). 1999. « La situation démolinguistique de la langue basque ». in : Blanchet (P.), Breton (R.) & Schiffman (H.), (ed.). 1999. *Les langues régionales de France : un état des lieux à la veille du XXIe siècle*. Louvain : Peeters, pp. 33-53.
- Weber (M.). 1983. *La fin des terroirs. La modernisation de la France rurale 1870-1914*. Paris : Fayard.