

## Les méthodes de la lexicographie dialectale : variation formelle et recueil des données

*Methods of Dialectal Lexicography: Formal Variation and Data Collection*

**Philippe Del Giudice**

---



### Édition électronique

URL : <http://journals.openedition.org/geolinguistique/1434>  
ISSN : 2650-8176

### Éditeur

UGA Éditions/Université Grenoble Alpes

### Édition imprimée

ISBN : 978-2-37747-166-9  
ISSN : 0761-9081

### Référence électronique

Philippe Del Giudice, « Les méthodes de la lexicographie dialectale : variation formelle et recueil des données », *Géolinguistique* [En ligne], 19 | 2019, mis en ligne le 09 décembre 2019, consulté le 19 décembre 2019. URL : <http://journals.openedition.org/geolinguistique/1434>

---

Ce document a été généré automatiquement le 19 décembre 2019.

Géolinguistique

---

# Les méthodes de la lexicographie dialectale : variation formelle et recueil des données

*Methods of Dialectal Lexicography: Formal Variation and Data Collection*

Philippe Del Giudice

---

## 1. Introduction

- 1 À certains égards, la description des dialectes s'avère plus complexe que celle des grandes langues institutionnelles. L'objectif de cet article est de faire le point sur les méthodes élaborées par les lexicographes dialectaux afin de faire face à deux défis particuliers : le traitement de la variation formelle et le recueil des données. Notre propos traite consécutivement de ces deux aspects et veut être avant tout une observation des pratiques en la matière, avec quelques prolongements théoriques.

## 2. La variation formelle

- 2 Le problème de la variation formelle se pose dans le temps, l'espace et la structuration de la société. Si les cartes commentées de Brun-Trigaud, Le Berre et Le Dû (2005 : 39-94) rappellent que la variation diatopique donne à voir différents stades d'évolution, on constate cependant que la variation diachronique n'a pas donné lieu à des traitements originaux dans les dictionnaires dialectaux. C'est donc sur les dispositions prises par les lexicographes pour rendre compte des variations spatiales et sociales que se portera notre attention, avant de se tourner vers d'autres aspects plus généraux.

## 2.1. La diatopie

- 3 L'espace dialectal que le dictionnaire retient peut varier « *from a hamlet to a continent* » Malkiel (1962 : 10)<sup>1</sup> en fonction des critères utilisés pour la sélection. Le dictionnaire s'intéresse-t-il à un ensemble linguistique ou à une variété ciblée ? Ce n'est pas la même chose de produire un dictionnaire du parler de La Brigue (village ligurien de l'arrière-pays niçois) comme celui de Lanteri (2006) ou un grand dictionnaire du berbère, chantier faramineux, peut-être irréalisable, qui attend encore les ouvriers qui oseront s'y attaquer. C'est toute la différence entre le dictionnaire monographique (isotopique ou microaéal) d'une pratique linguistique hyper-localisée et le dictionnaire diatopique, confronté à un tout autre niveau de diversité (cf. Rézeau, 1990: 1468-1469). Cependant, dans ce deuxième cas, le degré de diversité n'est pas forcément proportionnel à l'extension géographique de la langue considérée. L'exemple du petit département des Alpes-Maritimes le prouve. La variation que présentent ses dialectes romans primaires y est à tout point de vue plus spectaculaire que celle de l'ensemble des variétés régionales de l'anglais américain, dont les variantes lexicales et phonétiques ne remettent nullement en cause l'unité structurelle (phonologique, morphologique et syntaxique) générale de la langue. Une telle unité ne se retrouve pas dans le département dont Nice est le chef-lieu.
- 4 Encore faut-il délimiter la zone. Or les critères de délimitation sont multiples. Lorsqu'ils sont purement linguistiques, ils peuvent conduire à la réalisation de dictionnaires pandialectaux (comme le monumental LEI ou plusieurs dictionnaires catalans tels que le DCVB), que Barbato et Varvaro (2004 : 436) opposent aux dictionnaires régionaux. Ceux-ci peuvent également recouvrir, selon des critères linguistiques, un sous-espace dialectal homogène (à l'image du *Diccionari general de la llengua valenciana*<sup>2</sup> pour le catalan de la région de Valence ou de l'ALEC qui concerne le français populaire du Québec et de ses régions voisines) ; mais il n'est pas rare que la délimitation se fasse selon d'autres critères, sociaux, culturels, politiques ou géographiques. Le NALC pourvoit aussi bien des données relevant des variétés linguistiques du corse que des données liguriennes, en ce qu'il inclut (ce qui est naturel du point de vue géographique) le parler de Bonifaccio. Le DDDAM pour tenir compte des limites de la province historique du Dauphiné, mêle dans ses pages l'occitan (dans ses variétés alpines et non alpines déjà assez différenciées) et le francoprovençal. Le projet de l'ALPI était, quant à lui, prévu dès l'origine pour accueillir toutes les variétés romanes de la péninsule Ibérique. La délimitation de l'aire peut aussi être tributaire des conditions pratiques : facilité d'accès aux données ou ressources humaines disponibles. Le LMT, dont le principe général est de réunir le lexique oral de la moyenne et de la haute vallée de la Tinée, présente en outre les quelques données de la basse vallée qu'il était aisé de récupérer dans deux sources écrites antérieures (LMT : 11), et, plus étonnamment, un peu de vocabulaire de la Val Mairo, en domaine italien, où l'un des principaux informateurs du projet avait appris à tisser. Ne sont pourtant pas (ou très peu) représentées dans l'ouvrage d'autres communes tinéennes qui y auraient eu toute leur place. La conjoncture n'y a jamais rendu le collectage possible<sup>3</sup>. Dictionnaires pandialectaux ou régionaux d'un côté et dictionnaires monographiques de l'autre sont fort distincts d'un point de vue méthodologique. Le dictionnaire monographique ne pose pas de problème particulier. Lorsque toute une zone est explorée, en revanche, le lexicographe entre véritablement dans la dialectologie en embrassant la variation

spatiale caractéristique des dialectes. La nécessité d'intégrer des données diatopiques dans un objet lexicographique conduit alors à différentes stratégies de présentation.

- 5 (a) L'« approche dictionnaire » consiste à lister les principales variantes d'un terme. Chacune est alors porteuse d'une indication d'origine. C'est le choix du DDDAM où l'on trouve des entrées de ce type :

**Coucouei** [ku'kwɛj] (M) ; **coutouei** (DI) ; **coucouet, cougouet, coucot** (CH) ; **couet, coutet** (B) ; **coucouat** (A) ; **cutivè** (H) ; **coutiet, coutouit** s.m. 1. Derrière du cou, nuque, occiput. 2. Culbute, dans les Alpes.

Dans ce dictionnaire, les abréviations (M), (DI), (CH), (B), (A), (H) renvoient à l'origine géographique de la forme ou des formes qui les précède(nt), soit « moyen Dauphiné », « Diois », « Charpey », « bas Dauphiné », « Alpes dauphinoises », « haut Dauphiné ». Ces indications restent relativement imprécises. On ne sait pas dans quelles localités les formes ont été recueillies. *Coutiet* et *coutouit* ne sont, eux, pas localisés du tout. Lorsque l'origine concerne un sens, on s'aperçoit qu'elle est mentionnée en toutes lettres et qu'elle peut être très vague (« dans les Alpes »). Ce qui frappe également dans cet exemple, c'est la place que prennent les indications formelles par rapport au reste de l'information<sup>4</sup>. Malgré tout, ce procédé prend moins de place que des cartes. Toutes ces remarques seraient les mêmes pour la plupart des ouvrages qui ont opté pour l'« approche dictionnaire », y compris les plus reconnus (par exemple *Lou Tresor dóu Felibrige* ou TDF de Mistral).

- 6 (b) L'« approche atlas », parce qu'elle recourt aux cartes, est plus représentative de la géographie linguistique. Organisés onomasiologiquement, les mots dialectaux sont couchés sur la carte, chacun apparaissant au point d'enquête exact où il a été recueilli. Initiée par Gilliéron dans l'ALF, cette tradition se fonde sur le principe voulant que l'espace dialectal se reconnaisse à son homogénéité intrapunctuelle et sa variation interpunctuelle (cf. NALC : vol. 1, IV<sup>5</sup>). Potentiellement, l'atlas donne à voir de grands ensembles, cohérents lexicalement, phonétiquement, morphologiquement... Or, dans les faits, tout n'est pas si simple. Un même référent peut avoir plusieurs dénominations dans une même zone, et jusqu'en un même point (il s'agit alors d'une simple synonymie, pas toujours relevée ou signalée par les atlas). Cela ne va pas sans brouiller le message. D'autre part, il arrive que des commentaires de tous ordres soient nécessaires pour préciser la spécificité (sémantique, culturelle, morpho-syntaxique, etc.) d'une donnée ou les particularités de son collectage. En définitive, la simple exposition d'un lexème sur une carte se révèle rapidement insuffisante.
- 7 Mais tout représenter prend de la place. Partant, un arbitrage s'impose entre exhaustivité et lisibilité. À l'origine, les auteurs tranchaient en faveur de la lisibilité (c'est bien à cette intention que l'on doit le principe de l'ALF de ne retenir, autant que possible, que la première réponse de l'informateur). Peu à peu, on se mettra à redoubler d'ingéniosité pour tout faire apparaître sur la carte, en accumulant les transcriptions sur chaque point d'enquête, en développant les marges (l'ALMC et l'ALG font les deux), ou, pour la génération des atlas interprétatifs, en inventant des symboles (ALW), quitte à combiner plusieurs de ces procédés (à l'instar de l'ALAL), tout en ajoutant d'autres (recours à la couleur dans le PALDC<sup>6</sup> ou l'ALPA). Dans des travaux tels que l'ALE et l'ALiR, les commentaires organisés et systématiques qui accompagnent les cartes prennent la forme d'analyses motivationnelles si volumineuses qu'elles sont publiées dans des volumes distincts de ceux des cartes.

- 8 (c) L'approche mixte représente un compromis entre les deux voies précédentes. On l'illustrera par le DARE<sup>7</sup>, qui se veut dictionnaire, comme l'indique son titre (*Dictionary of American Regional English*), tout en produisant des cartes. Celle de la « guêpe maçonnerie » (cf. fig. 1) est, à elle seule, un exemple de classicisme d'atlas. On y voit notamment, à l'est, un clivage nord-sud (*mud wasp* vs *dirt dauber*) qui divise les treize états originels de la Déclaration d'indépendance, les deux principales zones étant plus ou moins bien garanties des influences mutuelles par la zone tampon de *mud dauber*.

Figure 1. – Carte de la guêpe maçonnerie dans le DARE.



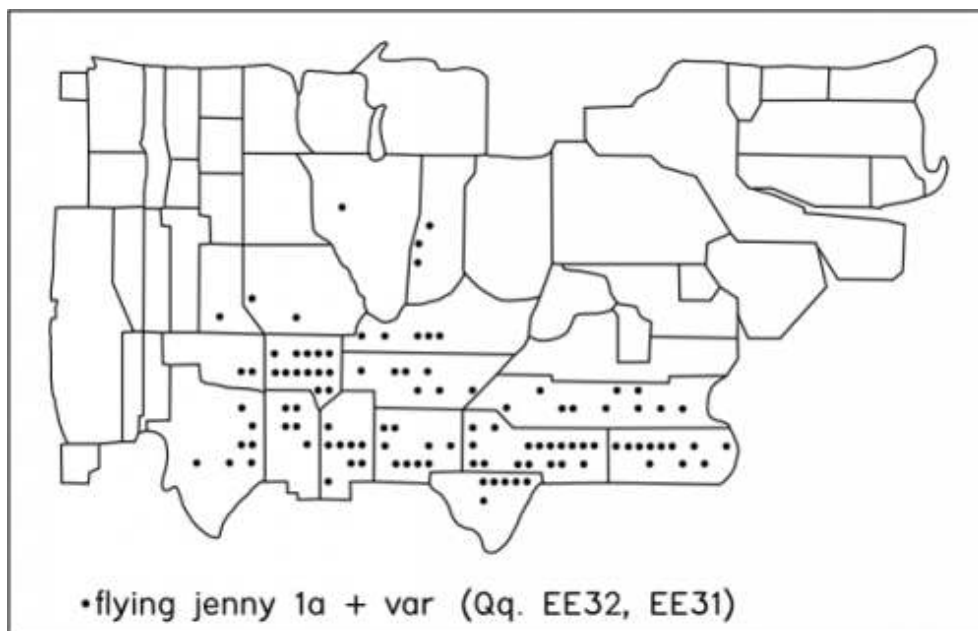
- 9 Le DARE n'est cependant pas seulement un atlas puisqu'il présente également ses données de manière sémasiologique sous la forme préconisée par l'« approche dictionnaire », en donnant des variantes graphiques, des indications géographiques (à l'aide d'abréviations diverses, comme dans l'article illustré *flying jenny* de la figure 2<sup>8</sup> où l'on peut lire « chiefly Sth » soit « principalement au sud »), les différents emplois, les synonymes (qui renvoient explicitement à d'autres entrées et, donc, à d'autres cartes), puis une somme d'attestations datées (des XIX<sup>e</sup> et XX<sup>e</sup> siècles essentiellement) et localisées, tirées le plus souvent de la presse locale ou d'enregistrements obtenus auprès d'informateurs par l'équipe de collectage du projet. Des indications étymologiques sont parfois aussi fournies. Ces lectures croisées reposent sur les renvois. Il y a notamment les renvois synonymiques, au sein des articles sémasiologiques (les articles de *mud wasp*, *mud dauber* et *dirt dauber* renvoient les uns aux autres) et les renvois à la section cartographiée du dernier volume, où se trouvent les représentations de synthèse comme celle de la guêpe maçonnerie.

**flying jenny** n Also *flying jinny*, ~ *jinnie*, ~ *ginny* [**jenny** 1, prob from its earliest form being a pole that was ridden astride (see 1a quot 1946)]

1 An amusement device in which riders are whirled in a horizontal circle; a merry-go-round; spec:a Any of various simple contrivances in which the riders supply the motive power. **chiefly Sth**, S Midl See Map Also called **flying board**, ~ **Dutchman**

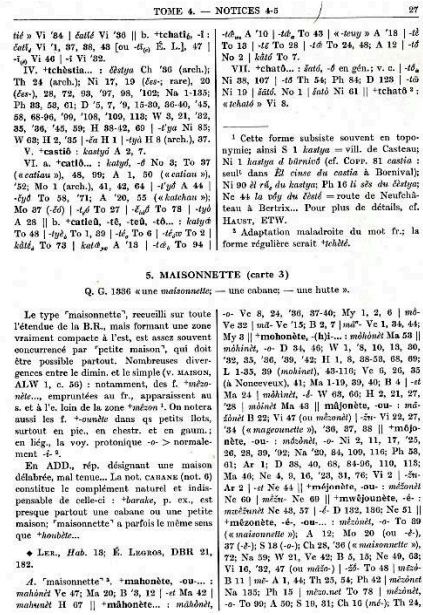
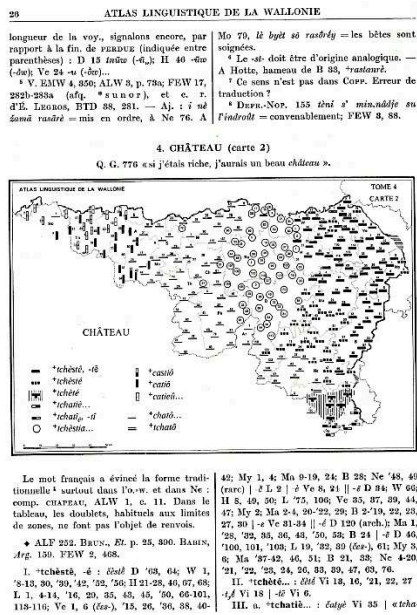
3, ~ **frappy**, ~ **horse 1b**, ~ **sally**, **jenny 6**, **spinning jenny**, **whirling** ~ 1876 in 1969 PADS 52.52 **neIL**, We had a spring board flying jinnie etc we had lots of fun. 1916 DN 4.268 **New Orleans LA, NC**, *Flying jinny*. . . A home-made form of carousel. [...]  
 2 See quot.1930 Shoemaker 1300 Words 23 **cPA Mts** (as of c1900), *Flying-ginny*— A small wind-mill, sometimes used at mountain communities to draw water or run a chop-mill.

Figure 2. – L'article *flying jenny* du DARE.



- 10 La carte<sup>9</sup>, placée à la fin de l'article, permet de localiser les attestations de l'unité lexicale.
- 11 L'approche mixte se retrouve dans des ouvrages qui préfèrent mettre en avant leur qualité d'atlas, comme l'ALW, dont, pourtant, la présentation rappelle davantage celle du dictionnaire. Effectivement, avec ses pages partagées verticalement en deux colonnes (à la manière des dictionnaires), avec ses petites cartes interprétatives qui suivent chacune une synthèse linguistico-historique fouillée, listant, dans des paragraphes, un foisonnement de variantes localisées (auxquelles s'ajoutent éventuellement des termes sémantiquement proches), l'ALW (fig. 3) n'a pas l'aspect d'un atlas linguistique classique :

Figure 3. – La présentation mixte selon l'ALW.

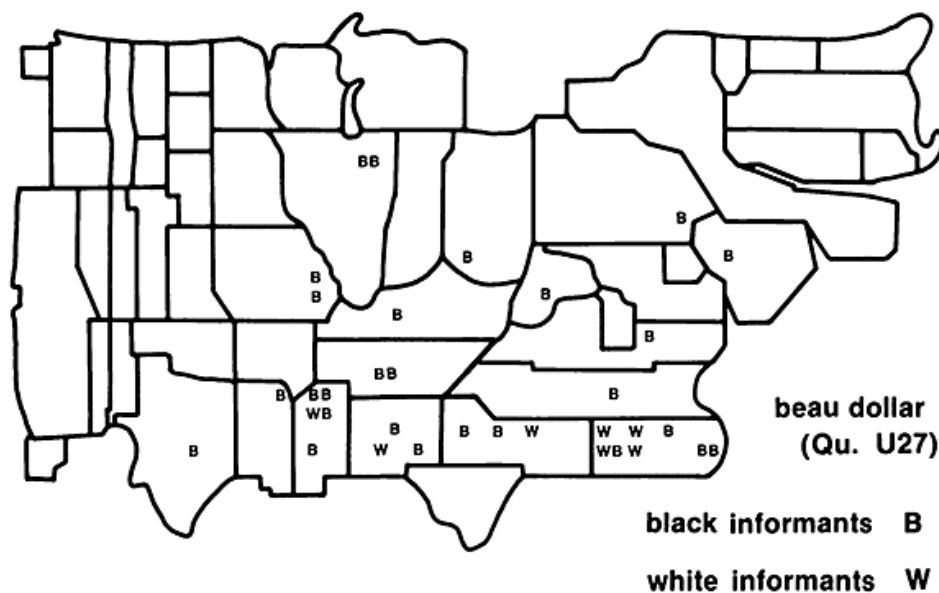


- 12 En représentant sur une même carte d'inspiration onomasiologique l'ensemble des lexèmes correspondants de la zone, l'ALW propose dans ses articles ce que le DARE réunit dans son dernier volume synthétique. Certains articles comportent en outre des précisions sémantiques minutieuses (cf. par exemple, l'article *Maisonnette*, réparti sur plusieurs pages). Lexicographie classique et géolinguistique sont donc susceptibles de cohabiter, suivant des modalités variables, pour transmettre une information diatopiquement différenciée. Les supports électroniques apportent en outre la possibilité d'accéder de manière optionnelle à des cartes affichées, voire produites, à la demande. C'est ce que fait un outil comme le THESOC.
- 2.2. Les sociolectes
- 13 Les dictionnaires dialectaux prennent bien peu en compte les variations liées à l'existence de groupes sociaux. Par conséquent, c'est avec une certaine peine que l'on trouve des ouvrages qui apportent quelque information à ce sujet. On peut y voir une conséquence inévitable de la diglossie et de l'abandon du dialecte aux seules classes inférieures dont découle la disparition de fait de la variation diastatique. On se dira peut-être aussi que les petites localités de campagne ne connaissent pas une diversification suffisamment importante de leur faible population pour que jaillissent ici et là de véritables traits sociolectaux. Ces croyances sont pourtant démenties pas les faits. À Saint-Étienne-de-Tinée, petite commune montagnarde d'un peu plus d'un millier d'habitants des Alpes-Maritimes, les habitants du bourg se piquent de pratiquer la variété haute (dont ils sont capables d'identifier les principaux traits) de leur parler qui les différencie du reste des habitants<sup>10</sup>.
- 14 À l'heure actuelle, le dialecte, souvent privé de norme prescriptive référentielle, souvent enseigné alors qu'il n'est plus transmis de manière naturelle, donne l'occasion de voir émerger des formes de variation sociale particulières. Il devient nécessaire de prendre en considération (a) l'écart entre la langue héritée des locuteurs natifs et la pratique des cercles de mainteneurs où la part des néolocuteurs croît sans cesse

(Bernissan, 2012; Martin, 2007 ; NALC : « Introduction ») ; et (b) la variation, au sein de la maintenance, que l'on doit à la coexistence de courants normatifs concurrents (Del Giudice, 2014). Il est donc fort regrettable que les dictionnaires dialectaux soient assez peu fournis en informations sur ces questions. On fera toutefois remarquer que le domaine d'oc n'est pas le plus mal loti. Le TDF précise que des termes sont propres à certains corps de métiers et, à Nice, le DLN signale les lexèmes recueillis auprès de la communauté juive, rappelant que jadis les nombreux juifs occitanophones possédaient des habitudes lexicales distinctives. Cependant, là encore, certains atlas vont plus loin en permettant l'accès aux métadonnées : des références telles que l'ALOC ou l'ALG présentent tous leurs témoins en introduction en précisant leur âge, leur sexe, leur lieu de naissance et leur profession. Ces présentations accueillent également diverses remarques. Pour chaque réponse apparaissant sur la carte il est possible de connaître le profil du témoin. Toutefois, aucune synthèse sociolectale n'est proposée dans ces outils. Il faut retourner dans l'univers anglo-saxon, et notamment dans l'Amérique du multiculturalisme, très friande de données sociales, pour s'apercevoir que les pratiques sociolectales se prêtent éventuellement à une représentation spatiale. Parce qu'elles combinent des critères sociaux et la localisation, certaines cartes du DARE sont un modèle de « sociogéolinguistique ».

- 15 Les critères du DARE sont de cinq ordres. Type de communauté (urbaine, montagnard, côtière), âge, niveau d'éducation, race et sexe. La figure 4 montre les attestations de *beau dollar* 'pièce d'argent' (pour *silver dollar*) en fonction de l'appartenance à la communauté noire ou blanche des informateurs.

Figure 4. – L'information sociolectale dans le DARE.



- 16 Il en ressort que sur les 34 informateurs qui ont proposé ce terme, 27 étaient des Noirs, alors même que les Noirs sont très minoritaires dans le réseau d'informateurs du DARE. Cette carte est en outre révélatrice d'un courant migratoire historique des Noirs vers les grandes agglomérations du Nord. Tous les témoins du Nord qui ont proposé ce terme étaient des Noirs qui vivaient à Chicago, Cincinnati, Philadelphie et Trenton. Pour Adams (2013 : 3) les caractéristiques du DARE en font une combinaison originale



de lexicographie historique et de géographie dialectale. On aurait envie de dire qu'il est également une combinaison réussie de lexicographie historique et de *sociogéographie* dialectale.

### 2.3. Autres problèmes généraux dus à la variation dialectale

- 17 Des types ciblés de variation ont permis de dégager deux démarches que l'on a appelées schématiquement la « démarche dictionnaire », fondée sur le paragraphe, les rubriques, les marques, les notices, et la « démarche atlas » qui recourt à la cartographie, les deux pouvant être associées pour produire des résultats concluants. Or, si l'on a vu qu'une représentation spatiale peut apporter des solutions intéressantes, toute la variation ne se laisse pas appréhender avec une égale efficacité par des cartes. Pour bien des données, le changement ne se rattache à aucune dimension. C'est notamment le cas de la variation non conditionnée. Ce n'est pas un hasard si la notion de polymorphisme, très proche à ses origines de la variation libre, a été élaborée puis explorée par des dialectologues<sup>11</sup>. La banalité de la variation dans l'univers dialectal suggère des modèles du changement linguistique qui réconcilient variabilité et systématisme. La lexicographie des langues minoritaires est en prise constante avec l'instabilité des faits. Le problème est le même pour la variation indéfinie (on constate une variation, mais on ne sait pas si elle suit un axe, et, si oui, lequel). Dans ce cas, seule la démarche « dictionnaire » est envisageable. Il faut alors se résoudre à donner les variantes sans indications explicatives.
- 18 Il arrive toutefois, même lorsque l'on connaît ses conditions et même lorsque l'on peut la cartographier, que la variation formelle ne puisse trouver toute sa place dans un projet de dictionnaire dialectal. Elle sera réduite ou totalement rejetée par le lexicographe. Les raisons peuvent être multiples, allant du souci de présentation à la volonté de ne donner que la « meilleure forme », par exemple celle du standard. Il y a donc trois cas de figure. (a) Le lexicographe cherche à retenir un maximum de variantes. Il s'agit alors de savoir comment il les organise. (b) Le lexicographe juge que la variation formelle est intéressante mais que la profusion sur laquelle elle débouche produit un message peu intelligible dans un dictionnaire ; il décide donc de n'en garder qu'un échantillon représentatif et les proportions de l'émondage seront variables. Dans ce cas, à la question de l'ordonnement s'ajoute celle des modalités du tri. (c) Le lexicographe rejette toute variation. Comment la forme retenue doit-elle alors être choisie ?
- 19 Ces trois cas montrent qu'il y a deux problèmes à traiter : l'organisation (cas a et b) et la sélection — modérée (cas b) ou extrême (cas c).
- 20 L'organisation a deux volets. Il y a d'abord celui de la concentration. Le lexicographe peut vouloir réunir des variantes proches, notamment purement phonétiques, sous une forme chapeau en graphie phonologisante dont elles sont mécaniquement déductibles. Les variantes sont donc implicitement prises en compte tout en limitant la taille de la liste fournie par le dictionnaire. Pour l'occitan, des formes telles que [tʃadj'ɛrɔ], [tʃadj'ɛβɔ], [sɑdj'ɛβɔ], [kadj'ɛrɔ], [kɑdj'ɛβɔ] se réduisent aisément par concentration aux formes graphiques *chadièro* et *cadièro*. Le deuxième volet est celui de l'ordonnement, c'est-à-dire l'ordre dans lequel sont rangées ces formes. S'il n'est pas aléatoire<sup>12</sup>, il peut répondre à divers critères (fréquence d'emploi, usure phonétique plus ou moins marquée, usage plus ou moins soutenu, etc.). Ces critères peuvent également être ceux

de la sélection. Les formes qui arrivent classées au-delà du nombre préalablement fixé de variantes par unité ne sont pas mentionnées dans le dictionnaire. Dans les projets les plus tolérants envers la variation, seules les occurrences les plus marginales disparaissent, dans d'autres c'est toute une masse d'occurrences, qui, mises bout à bout, représenteront éventuellement la majorité de l'usage, que l'on considérera comme superflues.

- 21 Dans l'ordonnement et dans la sélection, il y a hiérarchisation : toutes les variantes n'ont pas la même valeur. Il a été vu que pour ne pas surcharger son texte, le DDDAM donne la forme phonétique uniquement de la première variante. Quand elle est connue, c'est systématiquement celle du moyen dauphinois que l'auteur jugeait le plus représentatif de sa zone (Rixte, 2007 : 19).
- 22 Dans un autre ordre d'idée, même lorsqu'il ne s'agit pas d'« élaguer » le dialecte, l'identification d'une forme de référence est fondamentale, notamment dans les dictionnaires papier. La forme privilégiée fournit en effet le lemme d'entrée où trouver toutes les éventuelles variantes. Cette variété ne s'impose pas d'elle-même. Elle n'appartient pas forcément d'ailleurs au lexique du dialecte. On observe les solutions suivantes.

(a) Recours à une langue de référence hétéroglose. C'est le système adopté en France dans la série des atlas linguistiques par région où toutes les entrées sont en français normé, y compris pour l'Occitanie, la Catalogne française, l'Alsace...

(b) Recours à l'étymon. Le LEI, immense dictionnaire étymologique pandialectal de l'espace italo-roman, abrite tout le trésor lexical de l'Italie dialectale classé par étymons.

Outre le fait qu'elles ne feront jamais l'affaire pour les dictionnaires dont la langue d'accès aux articles doit être le dialecte<sup>13</sup>, ces deux options sont insuffisantes si le dictionnaire veut assumer sa mission totémique en présentant une autonomie dialectale. Il faut alors définir les formes de la langue objet (le dialecte) du dictionnaire. Pour cela, les lexicographes disposent d'autres solutions.

(c) Recours à un choix des unités au cas par cas, d'autorité, de ressenti ou au hasard de l'attestation. C'est un procédé courant des dictionnaires niçois.

(d) Recours à une sous-variété particulière (pour des raisons de prestige, de productivité, parce qu'elle a été la mieux décrite...). Il peut s'agir de la langue d'une localité précise, d'un grand auteur ou d'une école littéraire. Le PT, qui se veut dictionnaire de provençal, est surtout un dictionnaire de provençal rhodanien, très fidèle à l'exemple de Frédéric Mistral.

(e) Recours à une langue abstraite qui peut prendre au moins deux formes. (α) Le parler moyen, pratique idéalisée plus ou moins consensuelle d'une zone<sup>14</sup>. Cette notion, qui correspond au « *Sprachus* » (représentation intuitive de la « pratique normale » obtenue par comparaison des différents usages idiolectaux) de Hermann (1880 : 29), a ses limites lorsque les parlers sont assez différenciés. Il n'y a un « parler moyen » de la norme objective que pour des parlers relativement proches. Or, la délimitation de cette zone n'est pas une donnée, mais reste à déterminer : en fonction de celle-ci, le point d'équilibre du parler moyen se déplace. De plus, étant une construction abstraite reposant sur la conscience de locuteurs qui ne se sont pas concertés, il peut ne pas être parfaitement fixé, même lorsque son espace ne laisse guère de place à la discussion. (β) Une formalisation savante, construite dans le meilleur des cas sur la base d'une analyse de la variation de la langue. Cette formalisation s'inscrit le cas échéant dans un labeur plus vaste d'élaboration linguistique (CINLOA, 2008).

Encore faut-il, afin de considérer convenablement l'ensemble des faits, disposer de suffisamment de données. Or, elles ne sont pas toujours d'un accès facile.

### 3. Le recueil des données

- 23 Comment les lexicographes dialectaux diversifient-ils et exploitent-ils leurs sources ? Étapes préalables décisives, dont dépendront nomenclature et informations, la collecte et le traitement des données peuvent donner lieu à des stratégies élaborées, suivies de longues recherches érudites ou sur le terrain. Mais il n'est pas rare dans l'univers de la lexicographie dialectale que cette étape de réflexion sur les sources soit survolée. Les lexicographes dialectaux amateurs qui s'en exonèrent ne manquent pas, tels Jacques Guido, le « berger polyglotte », informateur du tendasque auprès des dialectologues de l'université de Nice et auteur entre autres d'un dictionnaire tendasque-français en cours de saisie. Pour réaliser son ouvrage, il s'est fié à ses seules compétences personnelles, sur le mode exclusif de l'introspection, avec les risques d'oublis qu'une telle méthode impliquait. Par conséquent, quoique précieux et certainement représentatif de la langue tendasque, son ouvrage pourrait être perçu comme étant le dictionnaire d'un idiolecte. Nul doute en effet que le lexique tendasque dépasse la connaissance d'un seul individu. Barbato et Varvaro (2004 : 433) mentionnent ce mode de fonctionnement en solitaire ou avec le concours de quelques amis. Pourtant, plusieurs approches sont envisageables pour constituer la base lexicale du dictionnaire. Les sources, de natures très variables, recouvrent des usages éventuellement peu destinés à la cohabitation. Il va de soi qu'une enquête de terrain effectuée en 2017 dans une zone rurale n'attestera pas tout à fait des mêmes usages qu'un dictionnaire vieillissant de la langue littéraire (quand elle existe). Le résultat d'un projet dépend donc en partie de la maîtrise des sources, ce qui nécessite de les identifier et, avec la perspective de construire le fonds lexical à exploiter, de les catégoriser. Il est généralement admis (Svensén, 2009 : 38) qu'il existe deux genres de sources : les sources primaires (locuteurs et textes authentiques) et les sources secondaires (descriptions de la langue, notamment dictionnaires et grammaires). Suite à l'étude des dictionnaires dialectaux de l'espace français, Rézeau (1990 : 1469-1470) estime que les méthodes de collecte de ces matériaux sont de trois types :
- a. La compilation des travaux précédents en un seul ouvrage. La plupart du temps, ce procédé entraîne des imprécisions de localisation.
  - b. La conduite d'enquêtes de terrain, auprès de témoins, avec la possibilité d'établir un réseau de correspondants.
  - c. La consultation de sources en tous genres (ce sont des textes, parfois anciens, pouvant être, pour l'onomastique par exemples, des cadastres, des actes judiciaires ou de notaires, etc.), qui peuvent conduire à la constitution d'un corpus mixte de sources synchroniques et diachroniques.
- 24 Dans ce classement, la méthode (a) correspond au recueil de sources secondaires, la méthode (b) aux sources primaires orales et (c) essentiellement aux sources primaires écrites ; ce qui s'accorde très bien avec la distinction binaire, tout en l'affinant un tant soit peu.
- 25 Pour observer la manière dont les dictionnaires dialectaux exploitent cette matière, on suivra l'ordre remanié (a), (c), (b) afin de passer progressivement des sources écrites aux sources orales, quoique ces dernières puissent présenter une forme hybride dont on discutera.

### 3.1. Le fonds lexicographique et le « dictionnaire de dictionnaires »

- 26 Le fonds lexicographique existant offre au lexicographe une réserve lexicale, parfois considérable, déjà triée et ordonnée. De plus, il fournit quantité d'informations utiles (définitions, catégories grammaticales, flexions...). En cela, rien d'étonnant à ce que la lexicographie d'une langue soit une construction historique, fondée pour partie sur une réutilisation permanente des travaux antérieurs. Dès l'introduction du *Grand Robert*, Paul Robert présentait d'ailleurs ouvertement la liste de tous les principaux dictionnaires français auxquels il empruntait (et qui avaient eux-mêmes pioché allègrement dans l'existant). La base d'un dictionnaire est parfois un dictionnaire unique. Celui-là est alors un genre de prolongement de celui-ci. Dans le domaine occitano-catalan, le DNFI copie le PVNI et ajoute des traductions françaises, comme le TDF part du DPF (Rostaing, 1974), pour ensuite étoffer la nomenclature et le contenu des entrées, modifier la forme ou la présentation, éventuellement éliminer des unités ou des informations jugées peu essentielles. Ainsi y a-t-il non seulement essai de compilation, mais aussi réévaluation des données. Avant l'ère numérique, le caractère systématique de la compilation était selon les cas plus ou moins abouti (dans la lexicographie niçoise, le DFN compile l'existant en commettant des oublis très regrettables). De nos jours la démarche, devenue courante, de numérisation et d'implémentation systématique des sources lexicographiques dans des bases de données offre de nouveaux avantages.
- 27 En effet, l'ensemble des unités lexicales qui, à un moment ou à un autre, sont apparues dans la lexicographie sont réunies sur un support informatisé et deviennent facilement accessibles suivant de multiples critères de recherche. Outre le fait que les carences de certains ouvrages sont complétées par les attestations des autres, l'origine multiple des données et des informations au sujet de ces données rend aisée la confrontation des descriptions. Dans cette phase de compilation, tous les dictionnaires, y compris les titres amateurs imparfaits, apportent leur pierre à l'édifice. Malgré leur indigence ou leurs approximations, ils peuvent toujours fournir un mot ou une expression attrapée au vol, un sens particulier, une interprétation éclairante, une indication géographique, des collocations, des exemples en discours (Van Keymeulen & De Tier, 2013 : 268).
- 28 Signe peut-être que la collecte de lexèmes dialectaux authentiques est entrée dans une phase de déclin structurel, la vogue, dans la pratique lexicographique dialectale, est à la production de *dictionnaires de dictionnaires*. De telles bases constituent alors bien souvent en elles-mêmes des produits finis. La réévaluation des données reculerait donc peut-être devant leur compilation pure et simple, même si ces outils peuvent s'avérer très innovants. Ainsi, la province du Trentin, avec l'ALTR, est dotée d'un dictionnaire (électronique) de trente dictionnaires, dont la plupart ont été produits après 1975. Il donne à voir, pour la première fois, le lexique de toute la province. En 2006, grâce à la mise en commun des dictionnaires locaux, l'ALTR comptait 35 000 entrées. Le projet de « *vocabolario di vocabolari* » (cf. Iannàccaro & Dell'Aquila, 2011) pour les parlers lombards, a la même démarche. Recouvrant toute la Lombardie à toutes les époques, son fonds lexicographique représente plus de quatre-vingts dictionnaires. À partir des formes données par les différents dictionnaires compilés, il génère toutes les variantes phonétiques d'un lexème, fournissant un grand nombre de variantes locales s'élevant parfois à plusieurs dizaines. Il donne aussi la possibilité de recréer les dictionnaires sources. Dès lors, la consultation se rapproche de celle du support papier original. La

banque de données des dialectes néerlandais présentée par Van Keymeulen et De Tier (2013) propose le même genre d'options. En redynamisant (grâce à des modes de recherche modernes, des nouvelles fonctionnalités, une présentation originale) des travaux antérieurs, pour certains inconnus ou introuvables, et en les resituant dans un grand spectre comparatif, ces dictionnaires de dictionnaires constituent une contribution de valeur à la connaissance des lexiques dialectaux.

- 29 La possibilité d'élaborer de grandes bases de données informatiques s'applique également à d'autres sources. Cependant, la stratégie qui consiste à constituer un corpus numérique de textes présente des caractéristiques fort différentes.

### 3.2. Le corpus de textes

- 30 Alors que la démarche « dictionnaire de dictionnaires » s'inscrit dans une logique de recyclage, la constitution d'un corpus de textes peut fournir la documentation utile à l'élaboration d'une nouvelle base lexicale véritablement originale. Celle-ci sera indépendante des précédents lexicographiques et pourra, si cela est jugé nécessaire, cibler un domaine particulier (soit qu'il ait été délaissé jusqu'alors par la lexicographie, soit qu'on décide de l'aborder avec une nouvelle documentation et des choix de traitement spécifiques).
- 31 Le TILG est un recueil informatisé de textes (plus de 3 000 documents répartis sur quatre siècles, des centaines d'auteurs et 30 millions de formes annotées<sup>15</sup>) conçu pour être la première étape de la formation d'un dictionnaire. Il doit sa genèse au fait que le galicien littéraire était très mal documenté. Effectivement, les dictionnaires du galicien sont presque tous « dialectaux », au sens microaréal, oral, populaire et fortement diglossique, ce qui a rendu nécessaire la fabrication d'un corpus. Cette stratégie vise donc à faire pénétrer dans la lexicographie des unités lexicales qui n'y ont jamais trouvé leur place. Projet de plus grande ampleur, le *Diccionari del Català Contemporani* en développement à l'Institut d'Estudis Catalans (IEC) depuis 1985 a pour objectif de recueillir le plus grand nombre possible de données sur l'ensemble du lexique catalan. Pour ce faire, le *Corpus textual informatitzat de la llengua catalana* a été constitué. Couplé à la base lexicographique BDLex, ce corpus lemmatisé s'étend sur 150 ans. Il regroupe 6 288 auteurs et 26 147 œuvres. Les textes, répartis entre écrits littéraires et non littéraires (presse, notamment), sont classés par auteur, titre, groupe chronologique décennal, groupe typologique, thème abordé, variété régionale employée... La deuxième phase du projet est l'élaboration du DDLC à partir du Corpus. La volonté déclarée de l'IEC de recueillir le plus grand nombre de genres textuels démontre que le grand dictionnaire de langue, véritable symphonie lexicale, doit couvrir tout un éventail d'usages. Cela dit, même le TILG, dont l'usage de référence est plus ciblé, cherche à être représentatif d'une certaine diversité et se montre soucieux d'inclure tous les types possibles de textes et de tenir compte, autant que faire se peut, de la diversité des profils d'auteurs et d'informateurs. Par rapport aux sources lexicographiques, le corpus donne une vision bien plus précise des usages en discours, et permet notamment de « *pinpoint the syntactic environment of the defined lexical form* » (Pereira & Gillier, 2012 : 269) en affichant la liste des occurrences contextualisées. Il faut toutefois s'attendre à des pratiques légèrement divergentes entre ce que l'on trouve à l'écrit et la langue telle qu'elle se manifeste dans la bouche des locuteurs (langue orale que l'on ne peut recueillir qu'à leur contact).

### 3.3. Les données de l'oralité

- 32 Le moyen *a priori* le plus direct d'obtenir des éléments de la parole dialectale est de les recueillir à la source, auprès des locuteurs. Sachant que le vocabulaire dialectal est essentiellement oral, « *the data are to be collected by way of fieldwork* » (Van Keymeulen & De Tier, 2013), ce qui permet en outre la localisation des données. Tous les atlas et de nombreux dictionnaires, à l'image de LMT, sont de bons exemples d'ouvrages lexicographiques produits sur la base d'un travail axé sur l'oralité. Il n'en reste pas moins que les procédures d'accès à ces données peuvent prendre plusieurs formes. Parler d'enquête lexicale de terrain comme d'un tout homogène serait hasardeux. Se pose la question de la méthode qui sous-tend l'enquête dont dépendront les données recueillies. Que s'agit-il de favoriser ?
- a. La traduction ? Elle s'applique à l'enquête lexicale brute ou peut s'articuler sur un questionnaire de phrases. La traduction permet de cibler les points de langue auxquels on s'intéresse, mais elle a le défaut de passer par la langue dominante, hors contexte, au risque de favoriser des calques plutôt que l'expression originale qui n'aura pas toujours un équivalent dans la langue dominante. Elle constitue la matière de la plupart des atlas linguistiques.
  - b. Le discours libre ? Le discours libre est probablement celui qui a le plus de valeur aux yeux de l'enquêteur soucieux de découvrir une langue authentique. Il est cependant chronophage lorsqu'il est question de traiter les données. Pour le lexique, s'il est l'occasion de faire de jolies trouvailles, il est loin de présenter l'efficacité d'un relevé. Les termes rares ou inconnus sont dilués dans la masse du vocabulaire hyperfréquent. Lorsque la recherche se focalise sur un thème ou une pratique culturelle, l'entretien dirigé (« racontez-moi comment on labourait, avec quels outils, comment on les construisait, etc. ») réduit considérablement cet inconvénient.
  - c. Le groupe de travail qui réunit des locuteurs<sup>16</sup> ? Lorsqu'il est possible de réunir suffisamment de témoins impliqués, le groupe de travail présente l'avantage de reposer sur le savoir de plusieurs natifs qui confrontent leurs connaissances et leurs représentations. La catégorisation du monde qui ressort des séances de discussion émane des membres de la communauté linguistique.
- 33 Si l'on souhaite avoir le regard le plus général sur la langue, l'idéal est de panacher ces méthodes afin d'accéder au lexique des locuteurs par diverses voies. Les conditions ne permettent cependant pas toujours de multiplier les enquêtes. Les enquêtes d'aujourd'hui viseront alors à compléter celles d'hier. Dans la dépendance vis-à-vis des travaux de terrain anciens, le problème du rapport entre oral et transcription est inévitable. Lorsque tout ce qui nous reste d'une enquête orale (non enregistrée) est une transcription graphique, s'agit-il encore d'oral ou d'écrit ? Ce n'est plus tout à fait la réalisation des témoins qui s'offre à nous, mais déjà une information traitée. Car on ne pioche pas sans difficulté dans l'oralité des lexies toutes faites, dépourvues d'ambiguïté. C'est toute la difficulté, classique, du rapport entre l'univers continu de la parole et les transcriptions discrètes. La matière brute, infiniment complexe et variable, glisse entre les doigts de l'observateur qui voudrait la saisir. Comment dès lors ne pas la trahir (l'interpréter) au moment d'en donner des représentations codées graphiquement ? Cette oralité que l'on n'entendra jamais ne se présente plus que sous une forme écrite. Comme l'a montré Dalbera-Stefanaggi (2001 : 23-24), la transcription (*a fortiori* la transcription en direct, sans enregistrement) impose de force une vision des choses,

interdisant d'autres interprétations potentiellement éclairantes pour l'étude de certains faits. En revanche, une production enregistrée mise à la disposition des chercheurs et accessible un nombre infini de fois offre une chance de mieux identifier et de mieux comprendre les points d'instabilité. Elle est également une alliée pour déceler les inévitables erreurs de transcription. Mais les enquêteurs n'ont pas toujours disposé comme aujourd'hui des moyens d'enregistrement. Les transcriptions sont alors des items de deuxième main, livrés par des « correspondants diachroniques ». En revanche, lorsqu'un enregistrement subsiste, reconsultable à l'infini, ces données de deuxième main redeviennent des données de première main. Tout cela ne réglera pas, de toute façon, le problème de la distance temporelle entre le moment de la confection du dictionnaire et celui des travaux antérieurs. Davantage que pour l'écrit, la parole évolue d'une génération à l'autre. Quel est le statut d'un item issu de l'oralité que les dernières générations ont abandonné et qui n'existe plus que dans une base surannée ?

- 34 La question inverse de l'écrit oralisé peut également se présenter. De quel domaine relève l'enregistrement d'un texte lu à la radio ? Effectivement, l'accès à l'oralité ne se fait pas uniquement par voie d'enquêtes. Une nouvelle dichotomie se présente à nous, au sein des données de première main, entre les fruits de l'enquête (qui consiste à aller chercher l'information) et le « tout-venant », qui par définition vient de lui-même à l'enquêteur. La recherche d'enregistrements historiques en tous genres participe à la connaissance d'un oral plus ou moins spontané. Ces sources prennent alors la forme d'émissions de radio, d'archives de la télévision, d'enregistrements familiaux, de disques. Pour bien des dialectes, elles sont assez peu nombreuses.
- 35 En définitive, des données intéressantes peuvent être tirées de l'oral, mais cette source n'est pas accessible sans difficultés. Encore n'a-t-il pas été question ici de la dimension humaine : il faudra que le témoin ait envie d'aider l'enquêteur (il pourrait être défiant), qu'il comprenne ce qu'on lui demande, qu'il ait de la mémoire, de l'intuition, une certaine sensibilité linguistique pour produire des énoncés intéressants, etc. Très importantes, les sources orales ne sont pas, à elles seules, la panacée du chercheur.

### 3.4. Dictionnaire général et nécessaire combinaison des types de sources

- 36 Si une observation devait ressortir de l'inventaire qui vient d'être dressé, c'est qu'en fonction de leur type, les sources (qui, prises séparément, ne documentent pas parfaitement le lexicographe) apportent des informations différentes et complémentaires, toutes indispensables à un grand dictionnaire général. Comment, en effet, avoir une vision large d'une langue si on ne va pas la chercher partout où elle laisse d'instructives traces ? Un ouvrage lexicographique ambitieux a tout intérêt à s'appuyer sur une grande base réunissant tous les types de sources. On a vu les mérites du dictionnaire de dictionnaires. Mais il ne faut pas perdre de vue que si la lexicographie se réduisait à la compilation, jamais aucun progrès ne serait possible. Le grand dictionnaire a naturellement vocation à dépasser ce stade. De la même manière que le dictionnaire de dictionnaires permet de mettre un terme à la dispersion des données lexicographiques, une grande base hétéroclite concentre l'ensemble des données lexicales disponibles, lexicographiques ou non, jusqu'alors inaccessibles ou éparpillées dans la nature. Elle joue un rôle de conservation, important dans la dimension muséale du dictionnaire. Le savoir lexical sur le portugais dialectal est

regroupé dans le *Tesouro Dialectal Português* (TEDIPOR)<sup>17</sup>, concentration de travaux monographiques, d'atlas, d'enquêtes linguistiques, de lexiques, de publications et de bases de données diverses. Le projet conserve intact toutes les définitions, toutes les descriptions apportées par les informateurs sur le terrain, toutes les tournures idiomatiques, et affirme sa double dimension, linguistique et culturelle (Pereira & Gillier, 2012 : 269). Sa vocation est de réunir un maximum de données lexicales, en explorant jusqu'aux sources les plus inconnues et les plus introuvables. Parce qu'il inclut de nombreuses sources hétérogènes, le TEDIPOR apporte un grand nombre de données neuves (pourtant recueillies de longue date), parfaitement absentes des dictionnaires portugais, ce qui a été démontré de manière spectaculaire à propos du champ lexical de l'ivresse (*ibid.* : 267-268).

- 37 Outre un volume de lexies plus important, la concentration de toutes les sources offre au chercheur les moyens d'améliorer la qualité de l'information en confrontant les attestations. Si le fonds lexicographique donne accès à une information synthétique, la base de textes permet de l'étalonner en la mettant à l'épreuve des faits (*ibid.* : 269). Inversement, le corpus de dictionnaires aidera à résoudre bien des problèmes pour rendre intelligibles des productions textuelles présentant des termes difficiles. Les données orales servent de la même façon à compléter les données attestées par l'écrit (on voit mal comment traiter le lexique et les prononciations d'une langue vivante si l'on n'a jamais été en contact avec la pratique des locuteurs) tout en étant parfois éclairées par elles. Cela revient à dire que chaque type de source permet de mieux saisir les données des autres.
- 38 L'écrit, souvent littéraire ou spécialisé, pas toujours très récent, ne donne qu'une faible indication sur la réalisation phonique des mots. Il est du reste difficile d'évaluer avec précision son degré d'artificialité, son décalage avec l'usage, sa tendance à l'hypercorrection, etc.<sup>18</sup>. En revanche, on trouvera dans l'oralité essentiellement du vocabulaire courant, mais à une époque où la diglossie a déjà fait son effet au détriment de certains usages (administratifs, scientifiques et de langue soutenue, notamment). L'oralité ne permet plus d'accéder à certains termes disparus — parfois à une époque très peu lointaine — oubliés ou remplacés par des emprunts. Sont évoqués dans l'ALW (IV, 9) des termes connus par des sources écrites « qui n'ont pas été relevés par l'Enquête. Dans certains cas, la liste est longue et elle pourrait encore être allongée sans beaucoup de difficultés ».
- 39 De ce point de vue, on peut regretter, lorsque la matière existe, que de grands dictionnaires, reposant uniquement sur un type de source, ne signalent pas leur parti pris dans leur titre (*dictionnaire général de l'oral de tel ou tel dialecte* ou *dictionnaire de la langue écrite de tel ou tel dialecte*). Le DDLC, véritable modèle de dictionnaire à la fois descriptif (il consigne aussi bien les « erreurs » que les variantes et les emplois diglossiques) et dialectal, ne se fonde que sur un corpus écrit, alors que le catalan, sous toutes ses formes régionales, est encore très vigoureux à l'oral. En matière d'exploitation de sources variées, l'ALW (IV, 9) montre que l'abondance d'attestations (enquêtes, lexicographie, textes, archives...) permet « de signaler des variantes, d'ajouter des localisations qui avaient échappé à l'enquête, de préciser le sens ou le genre d'un substantif, ou simplement de confirmer des données rares ou surprenantes », parfois même auparavant considérées comme suspectes. Van Keymeulen et De Tier croisent également les dictionnaires monographiques amateurs et les atlas, tant les premiers « *come as very welcome additions to the onomasiological*



*dictionaries [...], which are not capable of conveying semantic details and are altogether very poor in microstructure in general* » (2013 : 268).

- 40 C'est donc évidemment sur la base d'une documentation extraordinairement variée qu'ont été produits les monuments de la lexicographie dialectale. La documentation du GPSR est, à ce propos, impressionnante. Elle contient trois millions de fiches et plus d'un millier de manuscrits. Elle mêle le demi-million de fiches des correspondants produites au début du xx<sup>e</sup> siècle, un fichier toponymique de 150 000 noms de lieux, divers matériaux recueillis par les fondateurs du projet dans le cadre d'enquêtes de terrain, des fiches issues du dépouillement de différentes sources d'archives ou d'autres sources manuscrites ou imprimées, sans compter toute une collection de textes et de lexiques. Cette base n'a cessé de s'enrichir au fil des décennies. Elle contient, de plus, un large corpus iconographique de croquis, d'esquisses et de photographies, dont une enquête ethnographique de 1940 fournit la majeure partie. On trouve la même richesse dans le SI qui réunit la langue moderne des xix<sup>e</sup> et xx<sup>e</sup> siècles en exploitant les données apportées par un important réseau d'informateurs, des dictionnaires, l'écrit scientifique et littéraire, internet (et en particulier des *chats*), des atlas, des documents instructifs quant à l'onomastique régionale, des textes juridiques, des articles de journaux, des publicités... et la langue ancienne depuis le xiii<sup>e</sup> siècle, explorée via les nombreuses chroniques, les autres sources littéraires, les drames, les autobiographies, les journaux, les rapports de voyage, les légendes des saints, divers actes administratifs et juridiques, les sermons, les bibles, les dictionnaires, les pharmacopées, les correspondances privées, et quantité d'autres documents. D'autres dictionnaires, tels que le VPL, la FPB ou encore le DES, qui font coexister dans leur documentation des résultats d'enquêtes de terrain et des sources écrites remontant jusqu'au Moyen Âge, s'inscrivent dans la veine de travaux fondés sur des corpus complexes, panchroniques, dans lesquels tous les types de sources sont combinés. Tout grand dictionnaire général de référence du dialecte devrait donc veiller à ce que les différents types de sources soient exploités.
- 41 Reste à savoir comment articuler cette diversité. Selon la dimension, la facilité d'accès et de saisie/numérisation de ces différentes sources, selon aussi les caractéristiques du projet, il s'agira de définir un ordre de traitement, quitte à décider d'un type de sources de base sur lequel viendront s'adjoindre les sources de complément. On peut néanmoins faire quelques remarques générales.
- 42 Tout d'abord, le dictionnaire des dictionnaires, quand il est intégré à un projet qui l'englobe et le dépasse, a potentiellement deux fonctions. Pour l'utilisateur, il constitue une garantie par rapport à ce que lui apprennent les articles du dictionnaire, en ce qu'il donne à voir une partie de la matière brute à partir de laquelle ont été réalisés les articles du nouveau dictionnaire. En amont, il aura aidé le lexicographe à réaliser sa synthèse tout en faisant office de « garde-fou ». Chaque lexème bénéficie de la mise à contribution de l'ensemble de l'information lexicographique existante, observable en un coup d'œil. Cet exemple montre peut-être la voie pour les autres types de sources. Si, dans le cadre d'un projet électronique, les articles permettent au lecteur d'accéder à la base de textes, les mêmes facilités de contrôle (pour le lecteur) et la même attention rigoureuse (du côté du lexicographe) sont activées.
- 43 Avec ces considérations, on voit se dessiner une bipartition entre base de lexique brut et base de lexique en contexte. Les sources à classer peuvent alors être vues sous un jour nouveau, relativement indépendamment des catégories : sources primaires/

secondaires, sources écrites/orales, ou, s'il l'on reprend les distinctions (redondantes avec les autres mais structurées différemment) de Rézeau (1990), (a) dictionnaires compilés / (b) enquêtes / (c) écrits divers.

- 44 S'il n'y avait les données orales, la partition serait assez claire entre lexicographie (sources secondaires, (a)) et discours (sources primaires, (b)). Cependant, les données orales sont, de fait, lexicographiques (résultats des enquêtes lexicales, notamment par questionnaire) ou discursives, avec des zones d'entre-deux pour ce qui concerne le discours dirigé dont le statut sera à établir. Dans tous les cas, rien n'empêche d'opérer une bipartition des données orales sur le modèle des données écrites et de tout ramener à la classification *base lexicale vs base de discours*. On l'a vu, une fois transcrites, les enquêtes linguistiques sont ramenées à de l'écrit. Dans le dictionnaire des dictionnaires, les résultats d'enquêtes linguistiques lexicales ont parfaitement leur place et ne demandent aucun traitement spécial<sup>19</sup> : les données sont exploitables selon les mêmes modalités que leurs équivalents écrits. Le cas échéant, il suffit de préciser dans la base la date et les auteurs à la fois de l'enquête et de la transcription. Cela est aussi vrai pour le discours oral. Une fois transcrit, il devient un texte comme un autre. Seul son média d'origine (qui peut là encore être précisé dans la base) est différent. La bipartition des sources conduirait donc aux distinctions présentées par le tableau suivant.

	Sources écrites	Sources orales
Base lexicographique	Fonds lexicographique	Enquêtes lexicales
Discours	Corpus de textes	Discours libre, dialogues...

## 4. Conclusion

- 45 Le tableau auquel aboutit notre réflexion sur le traitement des données dialectales offre un modèle simplifié d'organisation des sources du dictionnaire. Ce n'est cependant qu'un socle théorique brut qui exige d'être couplé aux aspects évoqués dans la section 2 sur la variation formelle. Les questions d'état de langue et de variation diatopique, plus ou moins complexes selon le dialecte concerné et les délimitations effectuées, ne manqueront pas d'affecter cette base de départ qui reste tributaire de la richesse et de la nature des sources à exploiter. Cela revient à dire que les propriétés de l'espace dialectal à traiter ne sont pas neutres pour la lexicographie. Aucun projet de dictionnaire dialectal à la fois ambitieux et maîtrisé ne peut faire l'économie d'une étude linguistique variationnelle, parallèlement à l'inventaire des sources lexicales disponibles.

---

## BIBLIOGRAPHIE

### **Dictionnaires et atlas**

ALAL = BEYER Ernest (éd.), 1969, *Atlas linguistique et ethnographique de l'Alsace*, Paris, CNRS.

ALDC = VENY Joan & PONS I GRIERA Lúdia, 2001-, *Atles Lingüístic del Domini Català*, Barcelone, Institut d'Estudis Catalans.

ALE = ALINEI Mario *et al.* (éds), 1986-, *Atlas Linguarum Europae*, Assen-Maastricht, Van Gorcum / Rome, Istituto Poligrafico e Zecca dello Stato / Bucarest, Academia Română.

ALEC = DULONG Gaston & BERGERON Gaston, 1980, *Parlers populaires du Québec et de ses régions voisines (Atlas linguistique de l'est du Canada)*, Québec, Office de la langue française.

ALF = GILLIÉRON Jules & EDMOND Edmont, 1902-1910, *Atlas linguistique de la France*, Paris, Champion.

ALG = SÉGUY Jean *et al.* (éds), 1954-1966, *Atlas linguistique et ethnographique de la Gascogne*, Paris, CNRS.

ALiR = CONTINI Michel *et al.* (éds), 1996-, *Atlas linguistique roman*, Rome, Istituto Poligrafico e Zecca dello Stato.

ALLOC = RAVIER Xavier, 1978-1994, *Atlas linguistique et ethnographique du Languedoc occidental*, Paris, CNRS.

ALMC = NAUTON Pierre, 1955-1961, *Atlas linguistique et ethnographique du Massif central*, Paris, CNRS.

ALPA = BRUN-TRIGAUD Guylaine & LE DÙ Jean, 2011-2013, *Atlas linguistique des Petites Antilles*, Paris, CTHS.

ALPI = GARCÍA MOUTON Pilar (coord.), FERNÁNDEZ-ORDOÑEZ Inés, HEAP David *et al.*, 2016, ALPI-CSIC [[www.alpi.csic.es](http://www.alpi.csic.es)], édition digitale de Navarro Tomás Tomás (éd.), *Atlas Lingüístico de la Península Ibérica*, Madrid, Consejo Superior de Investigaciones Científicas.

ALW = REMACLE Louis, LEGROS Elísée, LECHANTEUR Jean, BOUTIER Marie-Guy *et al.*, 1953-, *Atlas linguistique de la Wallonie*, Liège, Université de Liège.

ALTR = CORDIN Patrizia (éd.), 2005, *Archivio lessicale dei dialetti trentini*, Trente, Università di Trento.

DAALF = ROBERT Paul, 1962-1965, *Dictionnaire alphabétique et analogique de la langue française : les mots et les associations d'idées*, Paris, Société du Nouveau Littré, Le Robert.

DARE = CASSIDY Frederic G. & HALL Joan H. (éds), 1985-, *Dictionary of American Regional English*, Cambridge, Belknap Press of Harvard University Press.

DCVB = ALCOVER Antoni M. & DE BORJA MOLL Francesc, 1930-1962, DCVB, *Diccionari català-valencià-balear: inventari lexicogràfic i etimològic de la llengua catalana*, Palma de Majorque, [Moll], 1975-1977 (édition mise à jour). Disponible en ligne sur <http://dcvb.iecat.net/>.

DDDAM = MOUTIER Louis, [1900], *Dictionnaire des dialectes dauphinois anciens et modernes*, Ms. [cf. Rixte (2007) pour le texte édité].

DDLC = RAFEL I FONTANALS Joaquim (éd.), *Diccionari descriptiu de la llengua catalana*, Institut d'Estudis Catalans.

DES = WAGNER Max Leopold, 1957-1964, *Dizionario etimologico sardo*, Heidelberg, Winter.

- DFN = CASTELLANA Georges, 1947, *Dictionnaire français-niçois*, Nice, Éditions ludographiques françaises.
- DLN = EYNAUDI Jules & CAPPATTI Louis, 1931-1939, *Dictionnaire de la langue niçoise*, Nice, Acadèmia Nissarda, 2009.
- DNFI = PELLEGRINI J. P. (Abbé), 1894, *Premier essai d'un dictionnaire niçois, français, italien absolument nouveau et inédit*, Nice, Imprimerie Robaudi Frères.
- DPF = HONNORAT Simon-Jude, 1846-1848, *Dictionnaire provençal-français ou dictionnaire de la langue d'oc, ancienne et moderne, suivi d'un vocabulaire français-provençal*, Digne, Repos.
- FPB = SELLA Alfonso, 1992, *Flora popolare biellese. Nomi dialettali, tradizioni e usi locali*, Alexandrie, Edizioni dell'Orso.
- GPSR = GAUCHAT Louis, JEANJAQUET Jules & TAPPOLET Ernest, 1924-, *Glossaire des patois de la Suisse romande*, Neuchâtel & Paris.
- LEI = PFISTER Max & SCHWEICKARD Wolfgang, 1979-, *Lessico etimologico italiano*, Wiesbaden, Dr. Ludwig Reichert Verlag.
- LMT = DOMENGE Jean-Luc, 2014, *Langue et mémoire de la Tinée*, Nice, Institut d'études niçoises.
- NALC = DALBERA-STEFANAGGI Marie-José, 1995-1999, *Nouvel Atlas linguistique de la Corse*, Paris, CNRS.
- PALDC = VENY Joan, 2007-, *Petit Atlas Lingüístic del Domini Català*, Barcelone, Institut d'Estudis Catalans.
- PT = FOURVIÈRES Xavier DE, 1902, *Lou pichot tresor, dictionnaire provençal-français & français-provençal*, Avignon, Roumanille.
- PVNI = CARLES Pietro, 1866, *Piccolo vocabolario Nizzardo-Italiano*, Ms.
- SI = STAUB Friedrich *et al.* (éds), 1885-, *Schweizerisches Idiotikon*, Frauenfeld.
- TDF = MISTRAL Frédéric, 1878-1886, *Lou Tresor dóu Felibrige ou Dictionnaire provençal-français*, Aix-en-Provence, Remondet-Aubin.
- TEDIPOR = DE AZEVEDO Maia Clarinda, VITORINO Gabriela, SANTOS Isabel, SARAMAGO João *et al.* (éds), 2010-, *Tesouro Dialectal Português*, Lisbonne, Centro de Linguística da Universidade de Lisboa.
- THESOC = DALBERA Jean-Philippe (éd.), 1992-, *Thesaurus occitan*, Université de Nice, laboratoire BCL (Bases, Corpus, Langage), UMR 7320.
- TILG = SANTAMARIA Antón (éd.), 1986-, *Tesouro informatizado da lingua galega*, Saint-Jacques-de-Compostelle, Instituto da Lingua Galega.
- VPL = PETRACCO SICARDI Giulia (éd.), 1985-1992, *Vocabolario delle parlate liguri*, Gênes, Consulta Ligure.

### **Autres références**

- ADAMS Michael, 2013, « DARE, Dialect, and Techniques of Historical Lexicography », dans R. W. Mc Conchie, T. Juvonen, M. Kaunisto *et al.* (éds), *Symposium on New Approaches in English Historical Lexis*, Somerville, MA, Cascadilla Proceedings Project.
- ALLIÈRES Jacques, 1954, « Un exemple de polymorphisme phonétique : le polymorphisme de l'-s implusif en gascon garonnais », *Via Domitia*, t. 1, fasc. 4, p. 69-103.

- BARBATO Marcello & VARVARO Alberto, 2004, « Dialect Dictionaries », *International Journal of Lexicography*, vol. 17, n° 4, p. 429-439.
- BERNISSAN Fabrice, 2012, « Combien de locuteurs compte l'occitan en 2012 ? », *Revue de linguistique romane*, vol. 76, n°s 303-304, p. 467-512.
- BRUN-TRIGAUD Guylaine, Le BERRE Yves & Le DÙ Jean, 2005, *Lectures de l'Atlas linguistique de la France par Gilliéron et Edmont. Du temps dans l'espace*, Paris, CTHS.
- CAMINADA Giulia, FIORONI Marco & GIRARDONI Francesca (éds), 2007, *Vocabolario del dialetto di Barni*, Côme, Provincia di Como.
- CINLOA (Comission internacionala per la normalizacion linguística de l'occitan alpin), 2008, *Nòrmas ortogràficas, chausias morfològicas e vocabulari de l'occitan alpin oriental*, Coni, Espaci Occitan & Regione Piemonte.
- DALBERA-STEFANAGGI Marie-José, 2001, *Essais de linguistique corse*, Ajaccio, Alain Piazzola.
- DEL GIUDICE Philippe, 2014, « Le dialecte niçois : usage consacré et parler de la rue », dans F. Neveu, P. Blumenthal, L. Hriba *et al.* (éds), *4<sup>e</sup> Congrès mondial de linguistique française*, Paris, EDP Sciences.
- HERMANN Paul, 1920 [1880], *Prinzipien der Sprachgeschichte*, Halle, Niemayer.
- IANNÀCCARO Gabriele & DELL'AQUILA Vittorio, 2011, *Lessico Dialettale della Lombardia - LDL. Dizionario generale delle parlate lombarde*, Relazione dei lavori al 15 marzo 2011.
- LANDAU Sidney, 2001, *Dictionaries. The Art and Craft of Lexicography*, Cambridge, Cambridge University Press.
- LANTERI Didier, 2006, *Dictionnaire français-brigasque*, Grasse, TAC-Motifs des régions.
- MALKIEL Yakov, 1962, « A Typological Classification of Dictionaries on the Basis of Distinctive Features », *International Journal of Lexicography*, p. 3-24.
- MARTIN Olitiana, 2007, *Étude de la langue occitane dans les écoles Calandretas de l'Hérault*, mémoire de recherche Master 2, Université Toulouse 2.
- PEREIRA Sandra & GILLIER Raissa, 2012, « TEDIPOR: Thesaurus of Dialectal Portuguese », dans R. Vatvedt Fjeld et J. Matilde Torjusen (éds), *15th EURALEX International Congress*, Oslo, Department of Linguistics and Scandinavian Studies, University of Oslo, p. 267-281.
- RÉZEAU Pierre, 1990, « Le dictionnaire dialectal : l'exemple français », dans F. J. Hausmann, O. Reichmann *et al.* (éds), *Wörterbücher : Ein internationales Handbuch zur Lexicographie / Dictionaries: An International Encyclopedia of Lexicography / Dictionnaires : Encyclopédie internationale de lexicographie*, Berlin & New York, de Gruyter, p. 1467-1475.
- RIXTE Jean-Claude, 2007, *Dictionnaire des dialectes dauphinois anciens et modernes par l'abbé Louis Moutier*, Montélimar & Grenoble, IEO Drôme & ELLUG.
- ROSTAING Charles, 1974, « Le dictionnaire d'Honorat source du *Tresor dóu Felibrige* », *Revue de linguistique romane*, vol. 38, p. 459-466.
- SIEPMANN Dirk, 2015, « Dictionaries and Spoken Language: A Corpus-Based Review of French Dictionaries », *International Journal of Lexicography*, vol. 28, n° 2, p. 139-168.
- SVENSÉN Bo, 2009, *A Handbook of Lexicography: The Theory and Practice of Dictionary-Making*, Cambridge & New York, Cambridge University Press.

VAN KEYMEULEN Jacques & DE TIER Veronique, 2013, « The Woordenbank van de Nederlandse Dialecten (Wordbase of Dutch Dialects) », dans I. Kosem, J. Kallas *et al.* (éds), *Electronic Lexicography in the 21st Century: Thinking Outside the Paper* (Ljubljana & Tallinn, 17-19 octobre 2013), Trojane, Institute for Applied Slovene Studies / Eesti Keele Instituut, p. 261-279.

## NOTES

1. Voir aussi Svensén (2009 : 35-36) selon qui les dictionnaires dialectaux « *may describe the dialect of a town or a parish, a part of a province or a whole province. There are also dialect dictionaries with a nationwide coverage (example: English Dialect Dictionary)* ».
2. Consultable en ligne (<<http://diccionari.lenguavalenciana.com/general/>>). Ce dictionnaire, qui se veut normatif, comme du reste le *Diccionari normatiu valencià* (<[www.avl.gva.es/lexicval/](http://www.avl.gva.es/lexicval/)>), n'intègre pas moins toutes les unités lexicales obtenues auprès d'informateurs dont les lieux de naissance et de résidence sont consciencieusement renseignés.
3. Pour l'ensemble de ces informations cf. LMT : 9-11.
4. Elle aurait pu être encore plus grande si, dans ce dictionnaire, chaque variante, et non seulement la première, était suivie de sa transcription phonétique.
5. « La nature même de la présentation de type géographie linguistique se prête mal à la collecte et à la restitution de variantes intra-ponctuelles : la conception d'un atlas présuppose la stabilité intra-ponctuelle et la variabilité inter-ponctuelle. La seule variation véritablement assumée est la variation géo-linguistique. »
6. Cet atlas constitue la partie interprétative de l'ALDC.
7. Les grandes caractéristiques du projet sont présentées par Adams (2013), Landau (2001 : 39) et le site internet du projet (<[www.daredictionary.com/page/howtoreadmaps/](http://www.daredictionary.com/page/howtoreadmaps/)>), dont sont tirés les cartes et commentaires présentés ci-après.
8. L'article original étant d'une longueur démesurée, il ne s'agit là que d'un échantillon arbitrairement tronqué.
9. Cette carte n'est pas au format conventionnel mais au format DARE, élaboré selon un calcul qui prend en compte la densité de population afin d'éviter les grandes aires vides du Midwest.
10. Selon le témoignage oral de Jean-Luc Domenge dont les travaux sont cités par ailleurs. Dans LMT (2014 : 27), l'auteur passe rapidement sur cette question. Un des informateurs enregistrés évoque aussi ces différences.
11. La définition d'Allières (1954) fait office de référence : « Nous appelons "polymorphisme" la coexistence, dans le langage d'un sujet parlant, de deux ou plusieurs variantes phonétiques ou morphologiques d'un même mot, utilisées concurremment pour exprimer le même concept, le choix de l'une ou de l'autre apparaissant comme indépendant du conditionnement articulatoire (tempo, etc.) ou d'une recherche quelconque d'expressivité. » Ainsi, on rencontre des phénomènes polymorphiques chez certains locuteurs français qui prononcent soit tantôt [swa], tantôt [swat] ou qui emploient indifféremment *je m'assieds* et *je m'assois*.
12. Un inventaire non ordonné est possible, même dans des travaux de grande qualité. Y a-t-il une progressivité logique des formes *empréurar, préurar, empreular, empresourar, empresurar, presurar, mete lou préul*, données dans cet ordre pour « emprésurer » (LMT : 276) ?
13. Ce problème se pose essentiellement au dictionnaire papier qui ne s'accommodera guère d'une foule d'entrées « creuses » ne servant que de renvois à la forme de référence (l'ampleur de la variation rendrait aussi le recours à un index — de proportions démesurées par rapport au reste de l'ouvrage — difficilement envisageable). Cet inconvénient est en revanche tout relatif dans un projet numérique qui accepte différentes formes d'accès à l'article.

14. (NALC : vol. 1, IV) : « Il semble que l'on puisse légitimement dégager, pour chaque aire étudiée — et c'est bien le concept d'*aire* qui s'avère opératoire —, une sorte de moyenne, de standard, fondé sur une norme plus ou moins explicite, ou plutôt "explicitable", tant par l'informateur que par le descripteur. Les études objectives des comportements langagiers tout comme les enquêtes menées sur la conscience linguistique, ou, tout simplement les discussions d'ordre métalinguistique avec certains informateurs concordent pour attester l'existence d'une variété reconnue par l'ensemble des locuteurs de l'aire considérée comme langue de référence, comme la manière "normale" de parler "ici" et "maintenant" (on a généralement conscience de l'"ailleurs" et de l'"avant"). Cette langue de référence remplit une fonction d'intégration symbolique à la communauté linguistique : on se reconnaît ou on s'affirme de tel ou tel *paese*. »

15. <<http://ilg.usc.es/TILG/>>.

16. Cf. l'expérience menée en Italie, qui a conduit à la publication du *Vocabolario del dialetto di Barni* (Caminada, Fioroni & Girardoni, 2007).

17. <[www.clul.ulisboa.pt/pt/\[...\]/677-tedipor-tesouro-dialectal-portugues](http://www.clul.ulisboa.pt/pt/[...]/677-tedipor-tesouro-dialectal-portugues)>.

18. Ne serait-ce qu'en français contemporain, Siepmann (2015) montre un écart confondant entre la langue des grands dictionnaires et la langue observée dans un corpus prenant en compte l'oralité.

19. Sauf l'éventuel aménagement d'un lien vers les enregistrements disponibles.

## RÉSUMÉS

Cet article offre un aperçu des solutions adoptées par les lexicographes pour faire face à des données dialectales insaisissables. Il apporte des réponses au questionnement suivant : quelles méthodes employer (a) pour représenter la variation géographique, sociale et libre dans les ouvrages de référence ? (b) pour recueillir des informations pertinentes et les structurer dans une base de données ?

This paper provides an overview of how lexicographers deal with elusive dialectal data. It gives answers to the following questions: what methods are used—or can be used— (a) to represent geographic, social and free variation in reference works? (b) to gather valuable information and arrange it in a database?

## INDEX

**Mots-clés** : lexicographie, dialecte, variation, recueil de données

**Keywords** : lexicography, dialect, variation, data collection

## AUTEUR

PHILIPPE DEL GIUDICE

Univ. Grenoble Alpes, CNRS, Grenoble INP\*, GIPSA-lab, 38000 Grenoble, France

\* Institute of Engineering Univ. Grenoble Alpes