

# Visualizing the *Dictionnaire des Régionalismes de France* (DRF)

Ada Wan

University of Zurich  
Zurich, Switzerland  
ada.wan@uzh.ch

## Abstract

This paper presents CorpusDRF, an open-source, digitized collection of regionalisms, their parts of speech and recognition rates, published in *Dictionnaire des Régionalismes de France* (DRF, “Dictionary of Regionalisms of France”) (Rézeau, 2001), enabling the visualization and analyses of the largest-scale study of French regionalisms in the 20<sup>th</sup> century using publicly available data. CorpusDRF was curated and checked manually against the entirety of the printed volume of more than 1000 pages. It contains all the entries in the DRF for which recognition rates in continental France were recorded from the surveys carried out from 1994 to 1996 and from 1999 to 2000. In this paper, in addition to introducing the corpus, we also offer some exploratory visualizations using an easy-to-use, freely available web application and compare the patterns in our analysis with that by (Goebel, 2005a) and (Goebel, 2007).

**Keywords:** digitization and release of heritage data, visual dialectometry, data visualization

## 1. Introduction

The DRF comprises data from the last large-scale study of lexical regionalisms in continental France in the 20<sup>th</sup> century, which took place from 1994 to 1996 and from 1999 to 2000. This paper describes a project carried out in 2016 with the following goals:

- i. to curate the data on recognition rates published in the DRF as CorpusDRF,
- ii. to provide preliminary analysis of the DRF data through hierarchical clustering, including exploratory visualization of some potential “regions” of lexical regionalisms based on the DRF, and
- iii. to evaluate the result from [ii.] by comparing with previous work performed by (Goebel, 2005a) and (Goebel, 2007) and by examination of the nature of our approach in relation to that of the survey design.

## 2. Data

The DRF is the first comprehensive book presenting a careful description of regionalisms of mainland France with up to 1,100 headword entries, furnished with definitions, examples, comments, citations and quotes, along with results from a survey of regionalisms carried out between 1994 and 1996 in France (EnqDRF), with a supplementary survey between 1999 and 2000, in form of recognition rates for the regionalisms in their relevant departments/regions as well as maps depicting the diatopic distribution of 330 regionalisms throughout France on the departmental level. Linguists from different regions compiled a list of known lexical facts characteristic of their regions which was then turned into a questionnaire – this resulted in about 4,500 facts being tested.

A few years ago, much of the DRF data fell victim to an administrative housekeeping effort. The corpus we curated manually from the printed version of this heritage volume is the only open-source digital version of the DRF recognition rates mapping entries (annotated with their parts of speech) with the corresponding departments.

### 2.1. Data documentation and profile

Entries (indicating a headword, a variant of a headword, or a multi-word expression involving the headword) with recognition rates published in the DRF were transcribed manually into a plain text file which was then processed into a data matrix in a TSV file with entries as rows and the 94 departments of mainland France as columns. (The department *Mayenne* was not included in the survey, and this did indeed result in an empty column in the matrix.) Most<sup>1</sup> entries in the DRF report a set of recognition rates for a certain number of departments/regions in France. Note that these departments/regions are pre-selected by specialists and that not the same set of words were surveyed in each department. Since our objective is to map entries by departments, names of regions such as *Champagne* and *Brittany*, had to be resolved into department names, leveraging cues from the legend and individual maps in the DRF wherever necessary (see Appendix A for the list of region-department correspondences used here). In the DRF, words or senses that were not surveyed are published with a “Ø”, whereas words or senses which were asked but not recognized by anybody have rates of “0” – we did not record the former (i.e. we represented such absence of data as empty cells in our matrix) and documented the latter with an actual “0” (zero). For some words, there are multiple sets of recognition rates recorded for their multiple senses, while for other words, sets of recognition rates for all the senses were merged and published as one set. Whenever there is a different set of departments for a distinct sense or expression, we tried to list it as a separate entry (with a sense number preceded by an underscore following the headword). In more complicated cases (e.g. a sense of a multi-sense headword having multiple senses/usages/variants), or cases where the merging of recognition rates seemed sensible, we merged the different rate sets and took the highest rate for each department reported. Words were surveyed based on the special senses they have taken on in a particular region,

---

<sup>1</sup>Please note that not all entries had recognition rates published in the DRF.

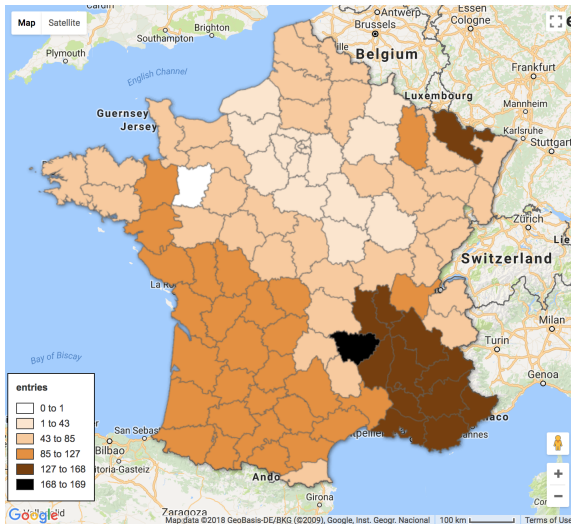


Figure 1: Entry distribution across departments – *Haute-Loire* (in black) has the highest number of entries at 168 while in *Mayenne* (white) no words were surveyed. (The intervals in the legend are half-closed with end values excluded.)

we therefore ask the users to refer to the DRF for further context.

All in all, we obtained 936 entries for CorpusDRF. Including the empty column for *Mayenne*, we have a data matrix with 936 rows and 94 columns. Of these 87984 ( $936 \times 94 = 87984$ ) cells, however, only 7248 were not empty, i.e. only 8.24%<sup>2</sup> of the matrix are filled. (Disregarding *Mayenne* would yield a matrix density rate of 8.33%.) The distribution of the entries with recognition rates is found to be rather unbalanced – more entries were asked in some departments than others (as shown in Fig. 1)<sup>3</sup>. For the 94 departments, we have a range of counts from 0 (*Mayenne*) to 168 (*Haute-Loire*) with a mean of 77.11, a median of 79.50, and a standard deviation of 38.07. (Discounting *Mayenne*, the figures remain similar with minimum at 5, maximum 168, mean 77.94, median 80, standard deviation 37.42.) The dichotomy between Northern/Central France and Southern/Southeastern France is clearly visible: there is almost a 50% gap between the two adjacent departments of *Allier* (count: 44) in the north and *Puy-de-Dôme* (80) in the south (the “southern department” with the lowest count on this “border”).

CorpusDRF will be made freely available without warranty on the ELRA/LREC website through the “Share Your LRs” initiative as well as through *Google Fusion Tables*<sup>4</sup> as “CorpusDRF”.

### 3. Related Work

Our approach is rather similar to most dialectometrical procedures. Dialectometry is a data-driven approach to iden-

<sup>2</sup>All numbers reported in this paper are rounded to 2 decimal places.

<sup>3</sup>See Appendix B for a list of these counts.

<sup>4</sup><https://sites.google.com/site/fusiontablestalks/home>

tify dialectal regions using statistical techniques on quantitative data. It is also sometimes known as “data-driven dialectology” or “quantitative dialectology”. Hans Goebl was a pioneering figure in this field as he was the first to use a computer for the calculation of linguistic distances and was hence able to process data on a bigger scale. He also introduced cluster analysis as a means of numerical taxonomy in dialectometry (Pickl and Rumpf, 2012). His collaboration with Edgar Haimlerl in the Dialectometry Project at the University of Salzburg since 1998 led to the development of the software *VisualDialectometry* (VDM)<sup>5</sup> which pioneered in the automatic combination of cartography and dialectometry. In the 2000s, visualization in linguistic analyses were also enhanced through the popularization of multidimensional scaling (MDS) by Wilbert Heeringa (Heeringa, 2004) and factor analysis by John Nerbonne (Nerbonne, 2006). Nerbonne et al. (2011) also released *Gabmap* as an online application for dialectology which allows researchers to inspect their data and visualize through MDS. For a more detailed overview of the historical development in (visual) dialectometry and a systematic overview in dialectology, please refer to Bauer (2009) and Boberg et al. (2018) respectively. For a comparison between VDM and *Gabmap*, please refer to Kellerhals (2014). In this paper, we report our methods and results in exploratory data analysis and visualization using *R* (R Core Team, 2015) and *Google Fusion Tables* (Gonzalez et al., 2010).

### 4. Data Visualization

*Google Fusion Tables* is a freely available web application that integrates seamlessly with *Google Maps*<sup>6</sup>. Producing a map visualization of data that contain geo information, in form of place names or Keyhole Markup Language (KML) data which define polygons on the map, is as easy as clicking a button. One can easily upload custom data files in tabular format and combine these with pre-existing public data resources. In our case, we imported our data matrix as a CSV file. To visualize the data in terms of their corresponding departments, we joined it with an already existing Fusion Table of KML shapes, made publicly available through *Fusion Tables* by GEOFLA<sup>7</sup>.

To view a map for an individual entry, select *Change feature styles* under *Feature map*, then under *Polygons/Fill color/Buckets*, select entry under *Column*. Fig. 2 shows how our map for the entry *verrine* compares to the one in the DRF in Fig. 3<sup>8</sup>.

We also uploaded the results of our clustering experiments as CSV files with departments as row names and their respective cluster IDs as columns. These files were then merged with the already prepared Fusion Table.

<sup>5</sup><http://ald.sbg.ac.at/dm/germ/VDM/>

<sup>6</sup><https://maps.google.com/>. All colored maps we created for this paper have been made using Map data ©2018 GeoBasis-DE/BKG (©2009), Google, Inst. Geogr. Nacional.

<sup>7</sup>[https://fusiontables.google.com/data?docid=1g\\_ydg74ooUSBzNfQBHOIdgrOKhxZD\\_92In8xTDg#rows:id=1](https://fusiontables.google.com/data?docid=1g_ydg74ooUSBzNfQBHOIdgrOKhxZD_92In8xTDg#rows:id=1)

<sup>8</sup>Figure taken from <http://www.atilf.fr/spip.php?rubrique86>.

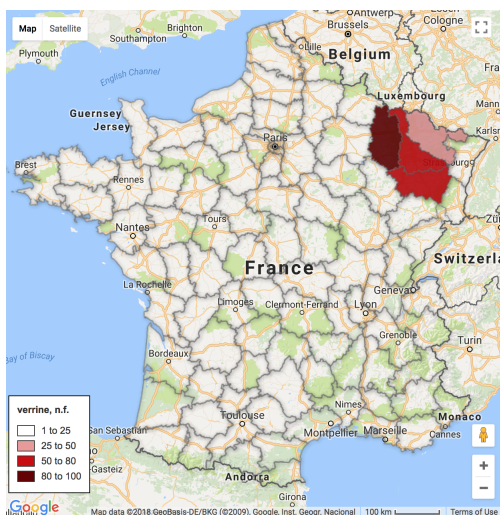


Figure 2: Our visualization of the entry *verrine*

△△ EnqDRF 1994-96. Taux de reconnaissance: Meuse, 80 %; Meurthe-et-Moselle, 75 %; Vosges, 70 %; Moselle, 25 %.



Figure 3: The entry *verrine* as printed in the DRF. (Note that “attestation” is only graphed in some of the maps in the DRF, not documented in text.)

## 5. Clustering

In order to detect potential regional groupings based on the recognition rates, we computed a distance matrix for our clustering with the continuous values we recorded as recognition rates by means of the Euclidean distance measure. The distance ( $d$ ) between points  $x$  and  $y$  in a Euclidean  $n$ -dimensional space is defined as follows:

$$d(x, y) = d(y, x) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

### 5.1. Treatment of empty values: NAs vs. 0s

In our data documentation phase, we made a distinction in coding “missing values” (i.e. the word was not asked in the survey for the department) as NAs as opposed to 0s (meaning 0% recognition rate). When NAs are treated as NAs, only words for which recognition rates are available for both will be considered in the distance/similarity com-

putation between two departments. This can be rather infeasible with our very sparse dataset – e.g., if 2 departments differ a lot in the distribution of their entries (i.e. which words got surveyed), but have 2 out of 5 common entries the same, then their dissimilarity score would be 0, whereas for 2 other departments with the same entry distribution, i.e. they were expected by the experts to share a common vocabulary of regionalisms, if all values of their recognition rates are different, even if slightly, they would have a larger dissimilarity score than the former pair. With a larger and denser dataset, disregarding a few empty values here and there may not be as problematic, yet not in our case. There is opportunity for leveraging various imputation techniques here, but we will leave that for future work. Instead we return to the context of the EnqDRF – if a word is not expected (by the specialists) to be known much or at all in a department, it is not surveyed there. This is an assumption that may or may not hold in reality and can only be proven with future empirical research with an extended survey.

In CorpusDRF, three versions of the data are provided – one with NAs left as empty, one with NAs imputed with 0s, and one imputed with -1s. The first two versions are for processing in *R*, the last is useful for an easy visualization in *Fusion Tables*.

### 5.2. Results and comparison with Goebel’s work on DRF

Hierarchical clustering algorithms were used as these are standard for taxonomical tasks similar to ours. Especially since we expected the distribution of our data to be a bit skewed, we performed hierarchical agglomerative clustering (HAC), i.e. we clustered agglomeratively (bottom-up), as opposed to divisively (top-down) so to minimize effect of the global distribution. This is also analogous to Goebel’s method. We first tried to replicate Goebel’s analysis of 4 clusters (Goebel, 2007) using Ward’s method with *R*’s built-in `hclust` function. Fig. 4 shows the analysis by Goebel while Fig. 5 shows ours with `hclust` (since `hclust` does not support missing values, we imputed the NAs as 0s).

A clustering algorithm that does handle missing data in *R* is Eisen’s algorithm, available via the `eisenCluster` method from the `hybridHclust` package. The resulting map using this algorithm is presented in Fig. 6. Note how the handling of NAs could have a huge impact on clustering results.

We also experimented with the unweighted average linkage (UPGMA) (Sokal and Michener, 1958). With UPGMA, the distance between two clusters is the average distance between all pairs of observations. UPGMA tends to join clusters with small variances and has slight bias toward producing clusters with the same variance (SAS Institute Inc., 2009) – this would be the preferred behavior for our data as we’d want the pairwise distances of all points to be small. Fig. 7 shows our results using UPGMA and `hclust`.

Goebel provided dialectometrical analysis of the DRF first by employing the 342 maps of individual word entries with recognition rates (2005a) and in (2007) the entire DRF data. The recognition rates presented in his analyses followed the format in the DRF: discretized into 6 bins. Although our methods differ – he used ordinal variables for recognition



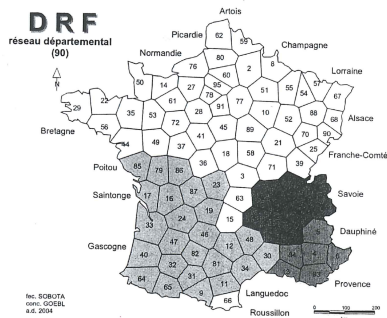


Figure 4: Visualization of 4 clusters from Goebel's analysis; map 22 from (Goebel, 2007).

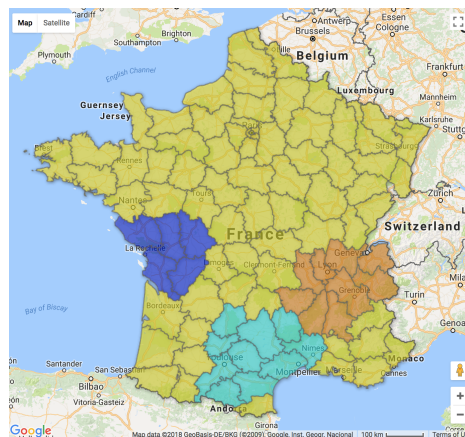


Figure 7: HAC with UPGMA (4 clusters) with NAs as 0s

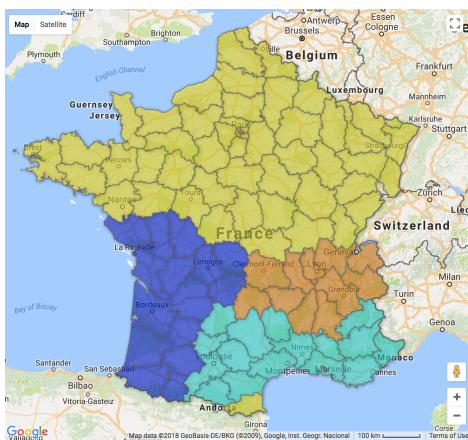


Figure 5: HAC with WARD (4 clusters) with NAs as 0s

rates and relative identity value (“Relativer Identitätswert”, RIW) as similarity measure, whereas we recorded recognition rates from the DRF text as continuous values and used Euclidean distance, there are parallels in our analyses. Fig. 7 exhibits overlap with his analysis (Fig. 4) in that the 9 departments in the Southeast (*Haute-Loire, Loire, Rhône, Ain, Haute-Savoie, Savoie, Isère, Drôme*, and

*Ardèche*) form a cluster of their own. This cluster differs from the area with denser entry distribution in the South-east in Fig. 1, suggesting that the pattern may not merely be contingent on the sheer counts of observations. Yet, the department *Pyrénées-Orientales* in southern France by the Spanish border shows the tendency to be clustered with the relatively homogenous northern half of France in all of our analyses shown here, mirroring its outlier status in entry distribution.

In general, we concur with Goebel's observation that the DRF data might be a bit too sparse and too unsystematic for proper dialectometrical analysis. But precisely because of this, we tried to map out some of these preliminary analyses so to inspire other researchers to further investigate.

## 6. Discussion

At the onset of this project, we had hoped to be able to identify French dialectal regions on the basis of the recognition rate data in a data-driven manner. That, however, did not pan out as expected not only due to data sparsity, but also because we have come to realize that it was not the intention of the editor/researchers of the DRF for the recognition rates to be evaluated quantitatively on a national level. No rigorous quantitative methods were applied in the survey design, as the main concern of the DRF was one of (qualitative) lexicography for humans – to investigate and compile a collection of regionalisms that were of questionable status in comparison to Standard French at the time. Entries with accompanying rate information were certainly asked, counts were tallied but the dataset was biased by the pre-conception of the experts who decided which words would be surveyed in which departments in the first place. Such a dataset could be more representative of the mental maps of the specialists than of the survey participants. That said, the publication of CorpusDRF could help effect an easy, clear, and systematic evaluation of the DRF. It could help us understand the perceptual dialectological situation of the researchers in the past so to design more comprehensive and quantitatively rigorous experiments and surveys to examine new regionalisms and to (re-)investigate the status and development of the old. (Many researchers in the past have advocated for more rigor and standard in experiment

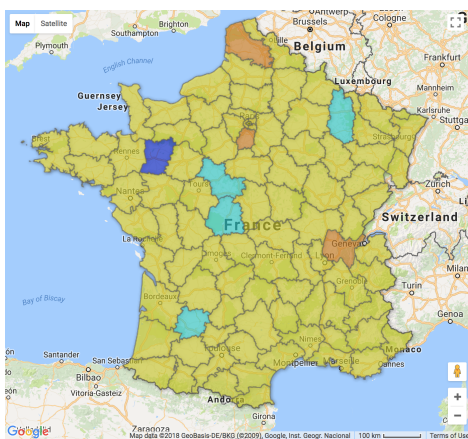


Figure 6: HAC using Eisen's algorithm (4 clusters) with NAs as NAs

design, readers may find sections on methodology and research design in Boberg et al. (2018), Schütze (2015), and Goebel (1993), among others, a worthy primer.) Data-driven methods can help us detect the hidden structures in the object(s) of investigation and, perhaps even more importantly, the hidden structures in the thought processes and values of the human investigators. That can in turn help much of (language) science progress in ways hitherto unanticipated and beyond a corpus-based approach of validation and refutation of hypotheses. Last but not least, CorpusDRF, in the absence of a digital version of DRF, can also function as a convenient reference facilitating studies in French lexicography, as it now enables an unprecedented production of maps for all and any of the 936 entries with recognition rate information.

## 7. Conclusion

In this paper, we described the DRF data profile and processing procedures and performed cartographic evaluation of some preliminary clustering results. Whereas it is clear that the nature of the EnqDRF and the DRF is different from that of a systematic dialectal survey and atlas (Goebel, 2005b), and that it would be presumptuous to generalize much from this dataset on a national level, it is nonetheless possible to gain insights on a smaller scale to better our understanding of the regions and of the language from the data as well as to use this as a stepping stone for future research on linguistic varieties and regionalisms. We hope that the accessibility of the CorpusDRF will inspire many more systematic and substantial surveys and dialectometrical analyses in the future.

## 8. Acknowledgements

We would like to thank Pierre Rézeau and l'ATILF (l'Analyse et Traitement Informatique de la Langue Française) for their support of the release of the curated data as a freely available language resource, as well as Hans Goebel, whose dedication in imbuing the field of dialectology and (Romance) linguistics with quantitative rigor and computational finesse (which we can imagine was a force a bit ahead of his time back in the days) shone through in our correspondences. Our thanks also go out to Mathieu Avanzi and Yves Scherrer for our valuable exchanges. All errors and misinterpretation remain solely the responsibility of the author.

## 9. Bibliographical References

- Bauer, R. (2009). *Dialektometrische Einsichten: Sprachklassifikatorische Oberflächenmuster und Tiefenstrukturen im lombardo-venedischen Dialektraum und in der Rätoromania*. Ladinia monografica. Istitut ladin Micurà de Rü.
- Boberg, C., Nerbonne, J., and Watt, D. (2018). *The Handbook of Dialectology*. Blackwell Handbooks in Linguistics. Wiley.
- Goebel, H. (1993). Die dialektale Gliederung Ladiniens aus der Sicht der Ladiner. Eine Pilotstudie zum Problem der geolinguistischen "Mental Maps". *Ladinia*, 17:59–95.
- Goebel, H. (2005a). Comparaison dialectométrique des structures de profondeur des cartes linguistiques du Dictionnaire des régionalismes de France (DRF) et de l'Atlas linguistique de la France (ALF). In M-D. Glessgen et al., editors, *La lexicographie différentielle du français et le Dictionnaire des régionalismes de France. Actes du Colloque en l'honneur de Pierre RÉZEAU*, pages 153–193. Presses Universitaires de Strasbourg, Strasbourg.
- Goebel, H. (2005b). L'interprétation de cartes dialectométriques (débat GOEBL-RÉZEAU). *Revue de linguistique romane*, 69:299–306.
- Goebel, H. (2007). La distribution spatiale des régionalismes du DRF comparée avec celle des données de l'ALF: un calibrage dialectométrique. In David Trotter, editor, *Actes du XXIVe Congrès International de Linguistique et Philologie Romanes (Aberystwyth 2004)*, volume I, pages 381–404. Niemeyer, Tübingen.
- Gonzalez, H., Halevy, A., Jensen, C., Langen, A., Madhavan, J., Shapley, R., and Shen, W. (2010). Google Fusion Tables: Data management, integration, and collaboration in the cloud. In *Proceedings of the ACM Symposium on Cloud Computing (SOCC)*.
- Heeringa, W. J. (2004). *Measuring dialect pronunciation differences using Levenshtein distance*. Ph.D. thesis, Rijksuniversiteit Groningen.
- Kellerhals, S. (2014). *Dialektometrische Analyse und Visualisierung von schweizerdeutschen Dialekten auf verschiedenen linguistischen Ebenen*. Master's thesis, University of Zurich.
- Nerbonne, J., Colen, R., Gooskens, C., Kleiweg, P., and Leinonen, T. (2011). Gabmap – a web application for dialectology. *Dialectologia: revista electrònica*, pages 65–89.
- Nerbonne, J. (2006). Identifying linguistic structure in aggregate comparison. *Literary and Linguistic Computing*, 21(4):463–475.
- Pickl, S. and Rumpf, J. (2012). Dialectometric concepts of space: Towards a variant-based dialectometry. In *Dialectological and folk dialectological concepts of space*, volume *Linguae and Litterae*, Band 17, pages 199–214. Walter de Gruyter, Berlin/New York.
- R Core Team, (2015). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Pierre Rézeau, editor. (2001). *Dictionnaire des Régionalismes de France*. De Boeck et Larcier s.a., Brussels, 1 edition.
- SAS Institute Inc. (2009). *SAS/STAT®9.2 User's Guide, Second Edition*. [https://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#statug\\_cluster\\_sect012.htm](https://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#statug_cluster_sect012.htm).
- Schütze, C. (2015). *The empirical base of linguistics: Grammaticality judgments and linguistic methodology*. Classics in Linguistics. Language Science Press.
- Sokal, R. R. and Michener, C. D. (1958). A statistical method for evaluating systematic relationships. *University of Kansas Science Bulletin*, 38:1409–1438.

## A Region-department correspondences used

The following list contains the correspondences between regional/departmental variants used in DRF → corresponding departments used in present work:

Alsace → Bas-Rhin, Haut-Rhin  
Aquitaine → Dordogne, Lot-et-Garonne, Gers, Hautes-Pyrénées, Pyrénées-Atlantiques, Landes, Gironde  
Argonne → Marne, Ardennes, Meuse  
Auvergne → Haute-Loire, Cantal, Puy-de-Dôme  
Basse-Normandie → Calvados, Orne, Manche  
Basse-Bretagne → Côtes-d'Armor, Morbihan, Finistère  
Bourgogne → Saône-et-Loire, Côte-d'Or, Yonne  
Bretagne → Ille-et-Vilaine, Côtes-d'Armor, Morbihan, Finistère, Loire-Atlantique  
Centre-Ouest → Deux-Sèvres, Vienne, Charente-Maritime, Charente  
Champagne → Aube, Haute-Marne, Marne, Ardennes  
Dordogne (nord) → Dordogne  
Franche-Comté → Haute-Saône, Territoire-de-Belfort, Doubs, Jura  
Haute-Bretagne → Ille-et-Vilaine, Loire-Atlantique, Morbihan, Côtes-d'Armor  
Haute-Loire (Velay) / Haute-Loire (nord-ouest) → Haute-Loire  
Haute-Saône (sud)/(nord)/nord → Haute-Saône  
Ile-de-France/Île-de-France → Paris, Seine-et-Marne, Yvelines, Essonne, Hauts-de-Seine, Seine-Saint-Denis, Val-de-Marne, Val-d'Oise  
Languedoc oriental → Gard, Hérault, Aude  
Languedoc occidental/Occidental → Ariège, Haute-Garonne, Tarn, Tarn-et-Garonne, Lot, Aveyron  
Limousin → Creuse, Corrèze, Haute-Vienne  
Loir-et-Cher sud / Loir-et-Cher (sud) → Loir-et-Cher  
Lorraine → Meuse, Meurthe-et-Moselle, Moselle, Vosges  
Meuse (nord) → Meuse  
Moselle (est) / Moselle (sauf est) / Moselle romane → Moselle  
Nord-Picardie → Somme, Aisne  
Normandie → Seine-Maritime, Eure, Calvados, Orne, Manche  
Picardie → Somme, Aisne, Oise  
Provence → Alpes-de-Haute-Provence, Alpes-Maritimes, Var, Bouches-du-Rhône, Vaucluse  
région lyonnaise → Rhône, Loire, Isère, Ain, Ardèche, Drôme, Saône-et-Loire  
Roussillon → Pyrénées-Orientales

## B Number of entries<sup>9</sup> by department sorted in ascending order:

Mayenne: 0; Eure: 5; Seine-Maritime: 5; Paris: 10; Val-de-Marne: 10; Yvelines: 10; Hauts-de-Seine: 10; Seine-Saint-Denis: 10; Essonne: 21; Eure-et-Loir: 27; Loiret: 27; Seine-et-Marne: 29; Val-d'Oise: 31; Nièvre: 34; Yonne: 36; Côte-d'Or: 36; Ardennes: 40; Marne: 42; Haute-Marne: 43; Indre: 43; Allier: 44; Aube: 44; Cher: 44;

Saône-et-Loire: 47; Indre-et-Loire: 50; Oise: 53; Pas-de-Calais: 54; Aisne: 54; Nord: 54; Somme: 55; Pyrénées-Orientales: 60; Loir-et-Cher: 63; Vosges: 64; Meurthe-et-Moselle: 66; Doubs: 68; Haute-Saône: 68; Calvados: 69; Jura: 69; Territoire-de-Belfort: 69; Orne: 69; Manche: 69; Finistère: 71; Côtes-d'Armor: 73; Morbihan: 73; Savoie: 73; Haute-Savoie: 76; Cantal: 79; Puy-de-Dôme: 80; Maine-et-Loire: 81; Haut-Rhin: 83; Sarthe: 83; Bas-Rhin: 83; Lozère: 84; Loire-Atlantique: 87; Ille-et-Vilaine: 88; Aude: 89; Gard: 90; Hérault: 90; Lot: 91; Meuse: 95; Haute-Vienne: 97; Corrèze: 98; Gers: 98; Creuse: 100; Vendée: 102; Dordogne: 103; Lot-et-Garonne: 104; Pyrénées-Atlantiques: 104; Hautes-Pyrénées: 104; Tarn-et-Garonne: 105; Charente-Maritime: 105; Vienne: 105; Landes: 105; Gironde: 106; Ariège: 106; Deux-Sèvres: 106; Charente: 106; Haute-Garonne: 106; Tarn: 107; Aveyron: 107; Ain: 121; Rhône: 129; Isère: 130; Drôme: 130; Ardèche: 131; Loire: 131; Hautes-Alpes: 134; Alpes-Maritimes: 136; Var: 137; Bouches-du-Rhône: 138; Vaucluse: 138; Alpes-de-Haute-Provence: 138; Moselle: 142; Haute-Loire: 168.

<sup>9</sup>entries with 0 included, with NA not