

5

Les modèles spatiaux et temporels dans la dialectologie

Comment séparer les tendances régionales et générales des fluctuations systématiques au niveau local

5.1. Rôle des modèles géographiques en dialectologie

Dans la mesure où les dialectologues font appel à la lecture comparative et typologique des cartes dialectales¹ ils tendent à trop accentuer les tendances régionales et globales au détriment des fluctuations systématiques au niveau local. Les variations locales et individuelles sont considérées comme des fluctuations fortuites et, par conséquent, d'un intérêt secondaire par rapport aux tendances plus globales.

Cette lecture typologique connaît également une variante temporelle, à savoir le principe selon lequel la répartition géographique des formes dialectales reflète leur évolution. Ce principe a été formulé de deux façons différentes. L'une, partant d'une division en deux classes, implique que, de deux formes, celle de la périphérie est plus ancienne que celle du centre (Bonfante, 1947). L'autre, partant d'une chaîne ordonnée, implique que la proximité géographique reflète la proximité chronologique, grâce à une chaîne de transitions allant de localité à localité (Weijnen, 1977; 3).² Nous reverrons un exemple de ce principe ci-dessous en discutant la théorie de la distance de la langue standard de Romaine (1986). L'interprétation selon laquelle la *géographie* des phénomènes dialectaux est le reflet direct du *temps* pose des problèmes, parce qu'elle ne rend pas compte de la différence entre les tendances de longue durée, les tendances de durée moyenne et les tendances de courte durée. En outre, le domaine géographique ne constitue pas une unité non plus: il faut distinguer les processus concernant la longue distance de ceux qui concernent les distances moyennes, ou ceux d'ordre local. Du reste, la tripartition parallèle des domaines temporel et spatial n'aura pas nécessairement pour conséquence que l'un est le reflet de l'autre (voir figure 1).³

¹ Voir Weijnen (1977), Hildebrandt (1983) en ce qui concerne les constellations géographiques générales (en allemand: *Kartenbild*).

² S'il est vrai que, parfois, ce principe est considéré comme un principe heuristique plutôt qu'un axiome, il est classifié souvent comme une loi (Weijnen (1977) et Bonfante (1947)).

³ Quant à la terminologie, la figure 1 ne représente pas une taxinomie et le même processus se caractérise parfois par les trois tendances à la fois. Au besoin, je me servirai des expressions

Les fluctuations individuelles et locales se caractérisent par certains aspects systématiques qui deviennent visibles après la soustraction des tendances plus globales. Grâce aux distinctions des dimensions temporelle et spatiale, que nous venons d'introduire, et tout en maintenant, sur le plan géographique, la distinction entre les tendances générales, les tendances régionales et les fluctuations locales (voir figure 1), il nous sera possible de prendre position par rapport à deux explications proposées récemment par rapport au problème de l'apocope du *t*-final dans les dialectes néerlandais.

Figure 1: Facteurs possibles d'un modèle explicite

| | Tendance générale | Tendance restreinte | Fluctuation syst./locale, individuelle | Fluctuation fortuite/locale, individuelle |
|----------------|-------------------|---------------------|--|---|
| Géographie | | | | |
| Temps | | | | |
| Domaine social | | | | |

5.2. Données et évolution

Goeman et Van Reenen (1985, ici chapitre 6) ont examiné le rôle que jouent les facteurs linguistiques dans le processus de l'apocope du *t*-final comme il se présente dans les dialectes néerlandais, par exemple: *knecht* > *knech* "servant", *hij loopt* > *hij loop* "il marche". Ils présentent également une analyse géographique provisoire de ces données pour arriver aux conclusions suivantes:

- a) il y a des changements très prononcés au sud-ouest de la Betuwe;
- b) le processus est en recul dans cette région;
- c) ce recul est peut-être en rapport avec la migration de l'ouest à l'est dans le passé;
- d) il y a eu toujours des alternances où le *t* n'a pas cessé d'exister: par exemple *knech* (sing.) à côté de *knechte* (plur.);

tendances générales, ou *processus à tendance générale*, etc. Le terme *facteur* est à concevoir comme une composante indépendante du modèle ou, plus naïvement, comme *facteur pertinent* (plus formellement: *variable indépendante*). Dans les modèles, les paramètres des facteurs indiquent l'ampleur (naïvement les effets) des tendances. Dans la figure, le domaine social n'a pas été réparti en facteurs.

e) l'expansion de l'apocope, et peut-être l'origine, doit être cherchée dans les régions où se sont rencontrés des dialectes apparentés, mais à caractère fort divergent;

f) le rôle de la langue standard est encore à préciser.

En analysant l'apocope du *t* (et du *d*) dans certains sociolectes et dialectes suédois, anglais et néerlandais, Romaine (1986, 614 s.) conclut qu'il faut distinguer deux processus:

a) l'apocope du *t*, qui serait un processus phonologique naturel;

b) le processus de réintroduction du *t* sous l'influence de la langue standard.

Moins la distance géographique entre le dialecte et la région où on parle la langue standard est grande, plus cette influence se ferait sentir (Romaine, 1986; 607-608).

En ce qui concerne les dialectes néerlandais, Romaine attribue à la règle de l'apocope du *t* le statut de marqueur sociolinguistique *stable*, c'est-à-dire que cette règle ne subit aucun changement en temps réel. Par contre, la règle de la *réintroduction* du *t*, causerait à un changement, elle engendrait une diminution de l'apocope. Un marqueur sociolinguistique se caractérise par la variation sur la dimension stylistique comme sur la dimension des classes sociales. La dimension stylistique est représentée, entre autres, par les données concernant les mots en contexte et les mots isolés.

Nous verrons plus tard (dans le tableau 1 ci-dessous) que les corrélations d'apocope du *t* pour les mots en contexte et les mots isolés sont significatives. Par conséquent, le comportement des mots en contexte est comparable à celui des mots isolés. Pourtant, la différence entre les corrélations est un indice de la stratification stylistique propre au marqueur sociolinguistique. Nous constatons que, au niveau des dialectes ce marqueur a un caractère spécial: la stratification en termes de classes sociales⁴ n'y apparaît pas.

Il y a quelques différences entre les explications de Romaine et de Goeman/Van Reenen qui méritent un examen détaillé. D'après Romaine, il n'y aurait pas d'expansion du processus de l'apocope. En outre, en ce qui concerne la diminution de l'apocope, elle adopte un point de vue comparable à ceux de Bonfante (1947) et Weijnen (1977): étant donné que le centre de la province de la Hollande Méridionale est à l'origine de la langue standard et qu'il en est resté le centre de gravité, moins les dialectes seraient éloignés, moins il y aurait des cas d'apocope. En revanche, l'explication de Goeman/Van Reenen laisse ouverte la possibilité de différences *locales* et n'exclut pas, à l'intérieur des dialectes, la possibilité d'évolutions déviantes de celles de la langue standard, comme l'expansion de l'apocope. L'explication de Romaine se limite aux tendances générales et globales, tandis que l'explication de Goeman/Van Reenen tient compte également des tendances locales.

⁴ Le test "ANOVA-deux facteurs à mesure répétée", appliqué aux mots en contexte et mots isolés pour déterminer l'effet de la classe sociale sur l'apocope, ne donne pas de résultat significatif ($F(4,45) = 1,417$, $p=0,2438$); l'effet de l'âge n'est pas significatif non plus ($F(3,45) = 1,882$, $p=0,1458$); l'analyse ne concerne que l'enquête de 1982.

Pour l'analyse du processus j'ai dépouillé deux enquêtes. L'une, datant de 1950, est publiée dans une série d'atlas régionaux, le RND. L'autre, commencée en 1982, est en cours.⁵ Van Hout (1980) a déjà analysé les données de l'enquête de 1950 concernant la Betuwe⁶ et Goeman (1986)⁷ a comparé les données des années 1950 et des années 1980 provenant de la même région.

Dans l'analyse statistique suivante nous comparerons une série de modèles du type "régression multiple". Par rapport aux publications antérieures (voir notes 5, 7) les résultats de notre analyse seront plus précis.⁸ Deux raisons ont déterminé le choix de ces modèles: 1) ils se laissent interpréter en termes de cause et effet (à condition que la direction des relations entre les variables soit été déterminée au préalable), 2) ils permettent non seulement de faire une distinction entre les diverses composantes postulées et leurs interactions, mais encore d'estimer l'importance des composantes. De cette façon, nous analyserons, en matière de l'apocope du *t*, comment la distribution régionale se sépare de la fluctuation locale, y compris les tendances parfois opposées.

5.3. La catégorie des modèles

Une localité, représentée par un locuteur dialectal, se caractérise par une proportion de réalisations dialectales Y_i . La localisation est déterminée par les coordonnées géographiques X_i . On retrouve ces éléments dans le modèle spatial:

$$Y_i = a + bX_i + e_i, \text{ où}$$

- b est la valeur numérique de la tendance générale, de l'influence des X_i sur les Y_i ,
- a est la moyenne générale,
- e_i représente les "erreurs": l'écart entre les Y_i estimés et les Y_i observés; l'écart comprend les erreurs de mesure des variables et / ou les variables pertinentes négligées.

Plus la mesure des coordonnées est précise, moins il y a des erreurs (cf. note 8). Chaque variable pertinente qui n'est pas incluse dans le modèle, ajoute à l'erreur

⁵ Il s'agit d'un projet commun de l'Université Libre d'Amsterdam et du P.J. Meertens-Instituut KNAW d'Amsterdam. Le but du projet est d'établir une banque de données dialectales. Voir: Goeman et Van Reenen (1985); Goeman (1987a).

⁶ RND=Reeks Nederlandse Dialectatlassen. Van Hout (1980) analyse le phénomène dans la région de la Betuwe en se servant de deux techniques: l'échelle d'implication et l'analyse de clusters. J'ai consulté les volumes 5: Zélande; 9: Brabant Septentrional; 10: Région des grands fleuves; 11: Hollande Méridionale et Utrecht; 12: Gueldre. J'ai dépouillé les données provenant des localités dans la mesure où celles-ci figurent aussi dans l'enquête de 1982. La date de 1950 forme plutôt la médiane des dates des enquêtes, qui ont eu lieu à partir des années trente jusqu'au milieu des années cinquante.

⁷ Cette étude traite le problème de la fiabilité des données des enquêtes dialectales traditionnelles.

⁸ Les coordonnées digitalisées dont je me sers ici ont été déterminées avec plus de précision que celles disponibles dans les études antérieures. La digitalisation des coordonnées locales a eu lieu d'après la carte de base de Kloeke (s.d.).

totale. Par conséquent, il est très important de ne pas omettre aucune variable pertinente.

Afin d'obtenir des estimations consistantes et efficaces, il faut que les variables aient une distribution normale, et qu'il n'y ait pas de corrélation entre les valeurs, pour mentionner seulement les deux conditions les plus importantes. Quand la distribution n'est pas normale, il suffit de transformer les valeurs pour assurer la normalité (voir note 14). Les répartitions géographiques tendent à se distribuer en régions cohérentes. Cela veut dire, que les concentrations régionales des phénomènes linguistiques, qu'elles soient positives ou négatives, impliquent *ipso facto* que les valeurs des phénomènes sont corrélées mutuellement et que les valeurs des variables sont autocorrélées. Ces autocorrélations sont importantes pour deux raisons, de nature complètement différente. D'une part l'existence d'autocorrélations est l'indice d'une cohésion régionale qui dépasse les tendances générales. D'autre part, par la présence d'autocorrélation, le test du modèle peut avoir comme conséquence le rejet du modèle. Pourtant, dans la mesure où le rejet est la conséquence d'autocorrélation, il est parfois prématuré. Si l'autocorrélation, tout en servant à dégager les tendances régionales et les fluctuations systématiques locales, déforme, en même temps, la tendance générale, il s'ensuit qu'elle déforme aussi la structure des erreurs e_i . Comme nous le verrons tout de suite, les erreurs peuvent servir comme l'indice des fluctuations systématiques locales. Quand le modèle est trop simple, il se peut que la répartition géographique se caractérise par des autocorrélations (voir le tableau 5 les *Données brutes observées*). Dans de tels cas, le choix d'un modèle non-linéaire peut faire disparaître l'autocorrélation des erreurs.

Pour rendre compte de la distribution géographique on peut se servir du modèle suivant:⁹

$$\begin{aligned}
 Y_i = & a + b_1 X_{i \text{ nord}} + b_2 X_{i \text{ ouest}} && \text{(partie linéaire basée sur les coordonnées nord et ouest)} \\
 & + (b_3 X_{i \text{ nord}}^2 + b_4 X_{i \text{ ouest}}^2 + b_5 X_{i \text{ nord}} * X_{i \text{ ouest}}) && \text{(expansion quadratique)} \\
 & + (b_6 X_{i \text{ nord}}^3 + b_7 X_{i \text{ ouest}}^3 + b_8 X_{i \text{ nord}} * X_{i \text{ ouest}}^2 + b_9 X_{i \text{ ouest}} * X_{i \text{ nord}}^2) && \text{(expansion cubique)} \\
 & + [g_i + e_i] && \text{(erreurs)}
 \end{aligned}$$

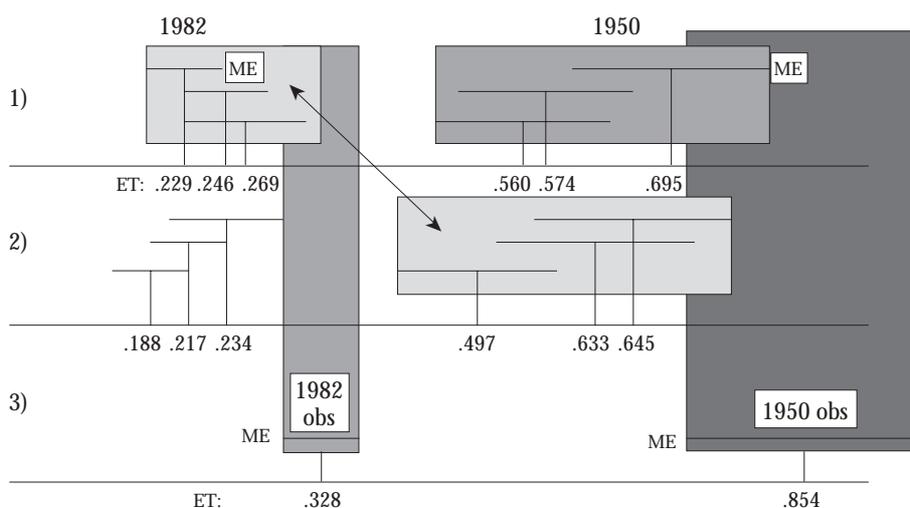
Dans ce modèle, nous distinguons deux sortes d'erreurs: e_i et g_i . e_i est la source des écarts fortuits, g_i est l'indice des écarts locaux systématiques. Il faut remarquer que $g_i + e_i$ se confondent souvent. Une solution pour isoler g_i s'appelle l'indice d'autocorrélation I de Moran (voir: Cliff et Ord, 1981).

Toutefois, si l'on accepte que les erreurs e_i , sources normales d'écart, sont minimales par rapport aux déviations d'ordre régional (ce qui se trouve être le cas, cf. note 8) on peut estimer les g_i en répétant l'expérience. Dans le cas présent, l'enquête de 1982 peut être considérée comme une répétition de celle de 1950.

⁹ Entre parenthèses: l'expansion polynomiale pour rendre compte des variations des tendances plus générales d'ordre nonlinéaire. Les produits sont les facteurs d'interaction.

Si, pour chaque localité, les écarts de la répétition sont centrés autour de la moyenne plus qu'autour de la tendance générale, g_i peut être conçu comme l'écart entre la tendance générale et la moyenne. Quand l'écart de la tendance générale (de l'enquête de 1950) est plus grand que l'intervalle de l'échantillon - exprimée par l'écart-type autour de la moyenne - on peut se faire une idée de cette déviation en tenant compte de la marge de l'écart-type (intervalle d'incertitude). La figure 2 montre la distance entre les écarts types, c'est-à-dire leurs déviations mutuelles. Pour les données des années 1950 et 1982, calculées séparément, les intervalles d'incertitude des écarts types des trois modèles se recouvrent (les barres noires dans la figure 2). En comparant les deux périodes, on constate que pour les trois modèles (linéaires, quadratiques et cubiques) les écarts types des données de 1950 sont beaucoup plus grands que ceux des données de 1982 et que les intervalles d'incertitude (1950 ~ 1982) ne se recouvrent pas (fig. 2.1). En effet, les erreurs de la répétition (1982: fig 2.1) sont centrées autour de leur moyenne au lieu d'être centrées autour de la tendance de 1950 (la prédiction: fig. 2.2). Par conséquent, il se peut qu'il soit question d'autocorrélation entre les données "brutes" (les observations), voir le tableau 5 ci-dessous. Pourtant, il n'est pas exclu non plus que le surplus de variance constaté soit dû à a) un facteur temporel, b) l'omission d'une ou plusieurs variables pertinentes, c) la nature des données. Nous discuterons ces trois possibilités une par une.

Figure 2.1-2.3: Comparaison des positions des écarts types (ET) de l'enquête de 1982 considérées comme répétition de l'enquête de 1950: 1) écart entre modèle et observations, 2) prédiction modèles (tendances), 3) observations. Ligne courte = modèle linéaire, moyenne = modèle quadratique, longue = modèle cubique. Barre noire: marge d'écart type (ME), région d'incertitude pour la position de l'écart type.



5.3.1. *La nature des données*

Les données de 1982 concernent pour la plupart des mots isolés, les données de 1950 surtout des mots en contexte. Pour chaque localité de l'enquête de 1982 nous disposons d'à peu près 160 mots. Quant à l'enquête de 1950, nous avons dépouillé de nouveau les 20 variables analysées par Van Hout (1980).¹⁰ Vu l'inégalité du nombre de mots des deux enquêtes, les proportions de 1982 sont plus stables que celles de 1950. Le fait que l'écart-type et la marge d'écart-type des données de 1950 sont plus grands que ceux des données de 1982 s'explique ainsi, au moins partiellement (voir fig. 2.3: données "brutes", observations). Nous constatons que la nature des données de l'enquête de 1950 et de celles de l'enquête de 1982 est essentiellement la même, parce que les corrélations entre les deux enquêtes sont significatives (voir tableau 1).

Tableau 1: Corrélations des proportions de l'apocope du -t entre *mots isolés* et *mots en contexte* pour 1950, 1982 et entre 1950 et 1982 pour mots isolés et mots en contexte

| | | mots isolés (a) | mots en contexte (b) |
|------|----------------------|-----------------|----------------------|
| 1950 | (a) mots isolés | – | .54 |
| | (b) mots en contexte | – | – |
| 1982 | (a) mots isolés | – | .33 |
| | 1982 | mots isolés (a) | mots en contexte (b) |
| 1950 | (a) mots isolés | .39 | .22 |
| | (b) mots en contexte | .58 | .30 |

5.3.2. *L'omission de variables*

Goeman (1987c, ici chapitre 4) a analysé les données de 1982 d'après le modèle adopté ici, complété de certaines variables de nature sociale: le sexe du sujet et de l'enquêteur et l'interaction entre les deux, l'âge, la classe sociale. En outre les variables: "distance géographique entre le dialecte et le centre historique de la langue standard" (cf. aussi Goeman, 1987a) et "perception de cette distance par les sujets parlants". L'hypothèse examinée dans cette étude est qu'un modèle tenant compte des facteurs géographiques (y compris la distance géographique de la langue standard) est suffisant et que les variables sociales ne comptent pas. Cette hypothèse s'est avérée correcte (voir tableau 2).

¹⁰ Les sujets de 1950 et de 1982 ne sont jamais les mêmes.

Tableau 2: Variance expliquée et t-valeur des variables indépendantes

| <i>Modèle étendu</i> | | |
|---|--------|----------|
| Backward: variable dépendante: t-apocope Les pas 1 à 11 incluent les variables indépendantes: <i>distance, sexe sujet, âge, status socio-économique, sudnordcub, sexe explor., ouestest, interactions sexe, distance perçue, sudnord, ouestestcub.</i> Les pas 12 à 17 éliminent dans cet ordre: <i>interactions sexe, âge, distance perçue, sexe sujet, status socio-économique, sexe explor.</i> Le résultat est le modèle restreint.* | | |
| <i>Modèle restreint</i> | | |
| <i>valeurs t-test des variables:</i> variable | t-test | sign. -t |
| distance | 2.845 | 0.0065 * |
| sudnordcub | -3.167 | 0.0027 * |
| ouest | 2.761 | 0.0082 * |
| sudnord | 2.157 | 0.0363 * |
| puestestcub | -2.457 | 0.0178 * |
| (constante) | -4.854 | 0.0000 * |
| Note: * Pour les équations avec les variables indirectes voir Goeman (1987c., ici chapitre 4). *=sign. | | |

Le modèle limité aux seuls facteurs géographiques suffit à expliquer 46% de la variance. Il n'y a aucune raison pour faire appel au modèle plus compliqué qui, avec toutes ces variables de caractère non-géographique, n'explique que 2% de plus (48%).

5.3.3. *Le facteur temporel*

Les années écoulées entre les enquêtes de 1950 et de 1982 auraient pu affecter les données: le comportement des sujets en matière d'apocope aurait pu devenir plus homogène ou plus hétérogène. Pour faire le test, j'ai intégré les données de l'enquête de 1950 comme variable dans la formule de l'enquête de 1982.¹¹ La variance expliquée augmente de façon significative pour les modèles cubique,

¹¹ Comme il suffit de se servir du modèle limité, il n'y a que deux équations à examiner: $Y_{1950} = a + X_{\text{nord}} + X_{\text{ouest}}$ et $Y_{1982} = a + Y_{1950} + X_{\text{nord}} + X_{\text{ouest}}$ (l'expansion polynomiale a été négligée pour simplifier l'exposé). Les deux équations ne sont pas simultanées mais récursives: elles contiennent les mêmes variables indépendantes exogènes (dans ce cas: les coordonnées géographiques). Comme ces variables sont indépendantes des écarts de 1950, les écarts de 1982 sont indépendants des valeurs de l'apocope pour 1950. On peut donc évaluer successivement ces deux équations (par "ordinary least squares" à condition que les écarts pour 1950 et 1982 ne covarient pas; cf. Pindyck et Rubinfeld, 322).

quadratique et linéaire.¹² La conclusion semble inévitable que la distance de 32 ans entre les deux enquêtes joue un rôle important, de sorte que l'enquête de 1982 ne peut être considérée comme une répétition de l'enquête de 1950. En outre, l'augmentation de la variance expliquée met l'accent une fois de plus sur l'équivalence des données des enquêtes de 1950 et de 1982. Étant donné que la possibilité d'estimer le facteur g_i par répétition est devenue minimale par le poids du facteur temporel, il nous reste à examiner, dans le paragraphe suivant, les écarts par rapport à l'autocorrélation géographique.

5.4. Trois modèles géographiques et leurs écarts

Dans ce qui précède nous avons constaté qu'il n'y a pas de variables pertinentes omises, que les données de 1950 sont comparables à celles de 1982 et que la dimension temporelle, elle aussi, joue un rôle important. Pour connaître la fluctuation systématique locale, il faut analyser l'enquête de 1950, l'enquête de 1982, et l'évolution qui s'est produite entre les deux.

Nous examinons les trois modèles - linéaire, quadratique et cubique - et leurs écarts,¹³ dans l'espace géographique. Le tableau 3 montre les coefficients des variables: les poids des facteurs du modèle cubique et du modèle quadratique et les valeurs du test-t qui indiquent si ces facteurs sont significatifs. Le tableau 4 présente l'analyse de la variance pour les trois modèles. Ajouter le facteur quadratique à l'enquête de 1950, et au modèle de l'évolution, a pour résultat une amélioration considérable de la variance expliquée (voir, dans le tableau 4, les colonnes "accrès"). Pour 1982, l'intégration de ce facteur est à peu près significative. L'intégration du facteur cubique ne fait guère augmenter le R^2 et n'atteint pas le seuil de la signification. C'est pourquoi j'ai présenté ici seules les figures tridimensionnelles des modèles quadratiques (voir fig. 3.1-3.2) seulement.¹⁴

Dans les modèles spatiaux, R^2 (=la mesure de variance expliquée) n'est pas le seul critère déterminant la question de savoir si le modèle est approprié. Pourtant, il est justifié de conclure que la variabilité est entièrement expliquée -malgré une valeur relativement basse de R^2 - à condition que les écarts ne manifestent pas d'autocorrélation. Par conséquent, l'intégration d'autres variables n'est pas nécessaire (voir Cliff and Ord 1981, 226). Le test de l'autocorrélation de Moran, I

¹² Respectivement $F=9.26$ (df 1,39) **, $F=6.25$ (df. 1,43) *, $F=13.833$ (df=1,46) ** (**: $p<0,001$; *: $p<0,05$).

¹³ Si l'on répète l'expérience en incluant les variables de caractère personnel - comme dans le modèle du tableau 2 - les résultats ne sont guère significatifs non plus.

¹⁴ Les coordonnées et les valeurs z s'expriment, dans ces figures, sur l'échelle de l'intervalle [0-1]. Les valeurs minimales sont l'équivalent de zéro et les valeurs maximales l'équivalent de 1. Les rapports géographiques entre les tendances de 1950 et de 1982 se laissent, de cette façon, visuellement comparer. Toutes les variables ont été recalculées sous la forme "écart de la moyenne". C'est pourquoi le paramètre a , la moyenne, ne figure pas dans les tableaux. Pour obtenir une distribution normale, les valeurs des variables dépendantes ont été transformées en flog (folded log), le "logarithme naturel plié" ($0.5 \log x$)-(0.5 $\log (1-x)$).

Figure 3.1: 1950 Tendence quadratique de l'apocope de -t

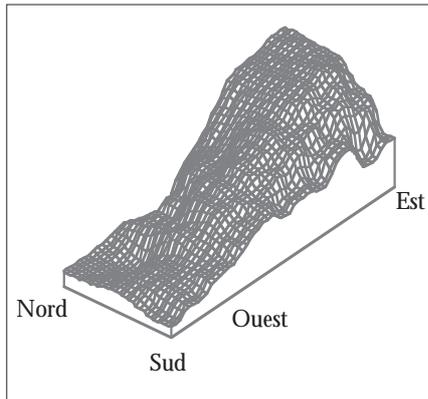


Figure 3.2: 1982 Tendence quadratique de l'apocope de -t

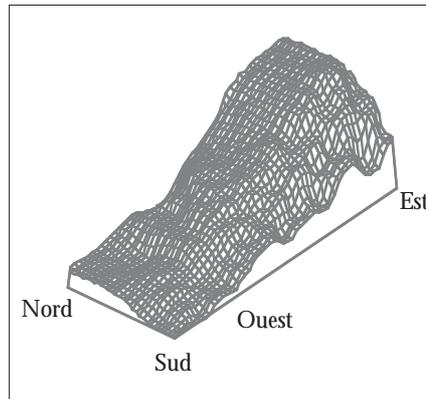


Tableau 3: Coefficients des variables des modèles cubiques en quadratiques. Variable dépendante: proportion d'apocope de -t

| Facteurs dans la régression | 1950 | | 1982 | | changement 1950-1982 | |
|--|---------------|-----------|---------------|-----------|----------------------|-----------|
| | β coeff | t-valeurs | β coeff | t-valeurs | β coeff | t-valeurs |
| ouestest | .374 | 1.278 | .909 | 2.915 * | -.519 | -3.860 * |
| sudnord | .249 | 1.114 | .214 | .899 | -.158 | -1.084 |
| ouestest ² | -.052 | -.376 | -.051 | -.347 | .181 | 1.488 |
| ouestest*sudnord | .424 | 2.554 * | .006 | .037 | -.253 | -1.779 † |
| sudnord ² | -.173 | -.657 | -.511 | -1.824 † | .308 | 2.178 * |
| ouestest ³ | .248 | .910 | -.061 | -.209 | | |
| ouestest ² *sudnord | .191 | .956 | -.003 | -.015 | | |
| ouestest*sudnord ² | .191 | .785 | -.566 | -2.175 * | | |
| sudnord ³ | -.050 | -.172 | .027 | .089 | | |
| d.d.l | | 40 | | 40 | | 44 |
| R ² | | .57 | | .51 | | .39 |
| Facteurs dans la régression | 1950 | | 1982 | | | |
| | β coeff | t-valeurs | β coeff | t-valeurs | | |
| ouestest | .665 | 5.650 ** | .614 | 4.743 ** | | |
| sudnord | .294 | 2.337 * | .433 | 3.079 | | |
| ouestest ² | -.145 | -1.383 | .003 | .280 | | |
| ouestest*sudnord | .288 | 2.345 * | .217 | 1.579 | | |
| sudnord ² | -.352 | -2.886 * | -.268 | -1.966 | à peu près sign | |
| d.d.l | | 44 | | 44 | | |
| R ² | | .55 | | .54 | | |
| **: sign. p < 0.001, *: p < 0.05 †: p < 0.08 | | | | | | |

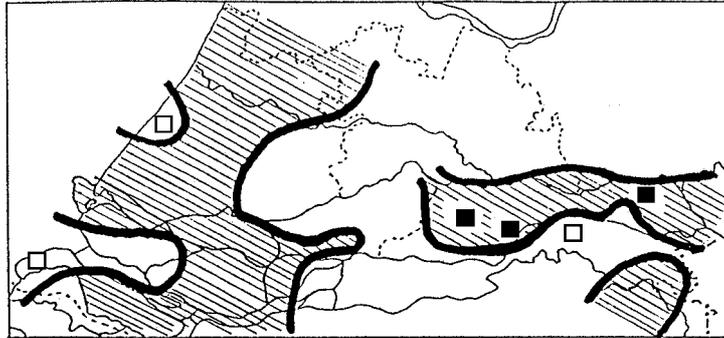
Tableau 4: Analyse de la variance et accrès de la variance expliquée

| | Source | d.d.l. | ss | F-test | Sign | Accrès de R ² | | Accrès expliqué | |
|---|----------|--------|--------|-----------------|------|--------------------------|--------|-----------------|--------|
| | | | | | | d.d.l. | F-test | d.d.l. | F-test |
| 1950 | Lineaire | 2 | 12.112 | (2,47) : 12.037 | ** | .34 | | | |
| | Quadr. | 5 | 19.693 | (5,44) : 10.723 | ** | .21 | 3,44 | 6.78 | * |
| | Cub. | 9 | 20.386 | (9,40) : 5.895 | ** | .02 | 4,40 | 0.491 | |
| | Ecart | 40 | 15.370 | | | | | | |
| 1982 | Lineaire | 2 | 1.727 | (2,47) : 11.463 | ** | .33 | | | |
| | Quadr. | 5 | 2.301 | (5,44) : 6.830 | ** | .11 | 3,44 | 2.733 | ‡ |
| | Cub. | 9 | 2.693 | (9,40) : 4.650 | ** | .07 | 4,40 | 1.633 | |
| Changement de 1950 à 1982 | | | | | | | | | |
| | Lineaire | 2 | 5.105 | (2,47) : 6.548 | ** | .22 | | | |
| | Quadr. | 5 | 2.301 | (5,44) : 5.750 | ** | .17 | 3,44 | 4.326 | * |
| | Cub. | 9 | 10.817 | (9,40) : 3.812 | ** | .07 | 4,40 | 1.218 | |
| **: sign. p < 0.006, *: p < 0.01, ‡: p ^a 0.05 (à peu près sign.) | | | | | | | | | |

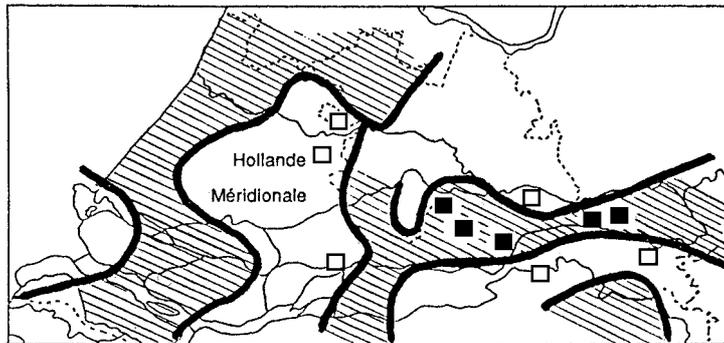
(Cliff and Ord 1981), est appliqué aux écarts en présupposant ou bien une distribution normale (N) ou bien une distribution inconnue (R: aléatoire). L'influence des localités entre elles est pondérée comme l'inverse de la distance carrée entre toutes les localités considérées deux à deux. La pondération rend compte de l'influence à très courte distance. Il est vrai que certains géographes affirment qu'on a pas le droit de se servir de cette mesure et qu'il faut intégrer cette distance dans le modèle comme variable pertinente. Pourtant, dans ce cas la seule variable concevable est la distance entre le dialecte et le centre géographique et historique de la langue standard. Comme cette variable n'est pas significative (voir ci-dessous), on n'a pas affaire à d'autres influences que celle d'une localité à l'autre. Etant donné que, à son tour, cette influence paraît diffuse elle peut être négligée dans le modèle. Le tableau 5 contient les résultats du test concernant les valeurs brutes, et les valeurs des écarts des modèles. Il en résulte qu'il y a autocorrélation parmi les valeurs brutes. Toutefois, l'examen des valeurs d'autocorrélation des écarts montre que les points géographiques ne sont pas autocorrélés. Cela veut dire que, dans les trois modèles, une partie satisfaisante de la variabilité est expliquée par les facteurs géographiques, que les composantes linéaires et quadratiques rendent compte de la plupart des effets généraux, et que la composante quadratique rend compte, en outre, d'une tendance de caractère régional.

S'il n'y a pas d'autocorrélation servant à mesurer les effets régionaux restreints, la seule possibilité de séparer g_i (la fluctuation locale systématique) de e_i (la fluctuation fortuite) est d'examiner les écarts de plus d'un écart-type. Les cartes 1, 2 et 3 montrent respectivement les écarts positifs ou négatifs de ces tendances générales. Il s'agit des données de 1950 (carte 1), des données de 1982 (carte 2), et de l'évolution du processus de 1950 à 1982 (carte 3). Les carrés sur les cartes indiquent les localités où les fluctuations dépassent systématiquement les fluctua-

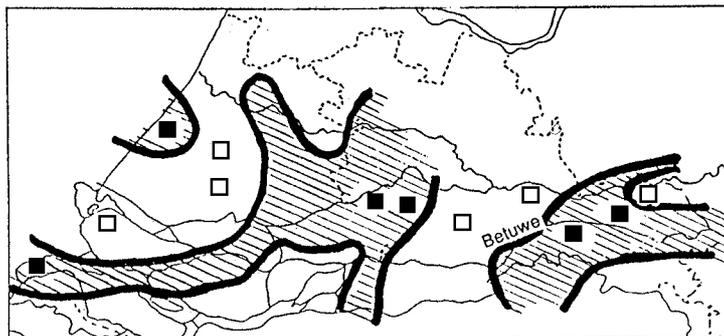
Carte 1: Ecart (résidu) de la tendance quadratique positifs et négatifs pour 1950



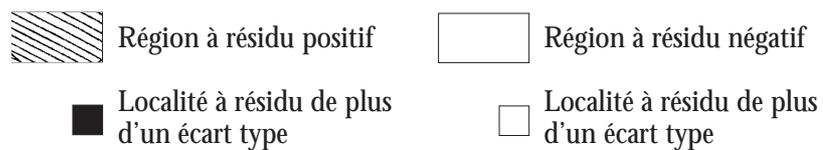
Carte 2: Ecart (résidu) de la tendance quadratique positifs et négatifs pour 1980



Carte 3: Ecart (résidu) de la tendance quadratique positifs et négatifs du développement de 1950 à 1980



Légende des cartes 1 - 3



tions positives ou négatives des autres localités, leurs écarts étant plus d'un écart-type. Les écarts des autres localités ne dépassent pas la valeur d'un écart-type, ces valeurs représentent en majorité des variations fortuites, en suivant les tendances générales et régionales (régions hachurées: résidus positifs; régions sans hachures: résidus négatifs). Pourtant, il y a des effets minimaux à courte distance parmi ces erreurs (cf. la pondération). Par conséquent, la différenciation constatée des régions représente donc sans aucun doute une ou plusieurs tendances réelles.

Quant aux tendances générales et régionales, l'apocope du *t* augmente de

Tableau 5: Test pour autocorrélation des écarts des modèles successifs et des observations

| | linéaire | Écarts 1950 quadratique | cubique | linéaire | Écarts 1982 quadratique | cubique |
|---------------|-----------------------------|----------------------------|---------|--------------------------|----------------------------|---------|
| I | .00449 | -.0128 | -.00023 | -.002445 | -.02748 | -.04618 |
| E(I) | -.02041 | -.02041 | -.02041 | -.02041 | -.02041 | -.02041 |
| z(N) | .70009 | .21362 | .56731 | .50506 | -.19889 | -.72482 |
| z(R) | .70113 | .21323 | .56704 | .50426 | -.19847 | -.72492 |
| | Écarts changement 1959-1984 | | | Données brutes observées | | |
| | linéaire | quadratique | cubique | 1950 | 1982 | |
| I | -.00867 | -.01868 | -.00298 | .15290 | .16572 | |
| E(I) | -.02041 | -.02041 | -.02041 | -.02041 | -.02041 | |
| z(N) | .33003 | .04857 | .48984 | 4.87278 ** | 5.23308 ** | |
| z(R) | .33171 | .04853 | .48868 | 4.93619 ** | 5.24307 ** | |
| **: p < 0.001 | | | | | | |

l'ouest à l'est et du sud au nord et les valeurs les plus élevées se trouvent dans le sud-est, comme le montrent également les facteurs (ouestest*sudnord) et (ouestest*sudnord²) de 1950. D'une part, l'évolution de 1950 à 1982 a abouti à une différence significative de l'ouest à l'est, de sorte que, plus on se trouve à l'est, plus l'apocope a *diminué*, d'autre part, il faut signaler une augmentation significative du sud au nord (sudnord²), à l'exception du sud-est (ouestest*sudnord). Les tendances plus globales ne confirment pas l'affirmation, de Romaine, que moins on est éloigné du centre de la langue standard, plus la diminution de l'apocope est forte.

Quant aux tendances locales, un changement important se produit dans l'ouest de la Betuwe, où l'apocope a diminué, ainsi que dans le centre de la Hollande (carte 3). On serait enclin à attribuer ce changement, encore une fois, à l'influence de la langue standard comme l'ont fait Romaine (1986, 607-608), et par rapport aux formes verbales, Goeman (1987a). Si, afin de déterminer le rôle de cette influence, on inclut dans le modèle la variable *distance*, représentant l'influence de la langue standard (l'intégration de cette variable pourrait encore

représenter une partie de la variabilité des *e*) la variance supplémentaire expliquée n'est significative ni pour les données de 1950, ni pour celles de 1982, ni pour l'évolution de 1950 à 1982.¹⁵ Si le changement de 1950 à 1982 n'est pas dû à l'influence *externe* de la langue standard il faut l'attribuer, en majeure partie, à l'évolution autonome et *interne* des dialectes par rapport à l'apocope. Pourtant il n'est pas exclu qu'il soit question d'une influence plus diffuse de la langue standard, sans rapport avec la distance géographique.¹⁶ Les dialectes limitrophes de la région hollandaise - comme les dialectes occidentaux de la Betuwe - se caractérisent par un renforcement remarquable du processus de l'apocope du *t*. Ces dialectes sont sujets, d'ailleurs, à des tendances fortement contradictoires (voir cartes 1, 2 et 3).

Projetés dans l'espace les écarts montrent, sur les trois cartes, un mouvement ondulant du nord-ouest au sud-est.¹⁷ Par conséquent, les tendances générales, la distribution régionale et les fluctuations systématiques locales se laissent distinguer sans se séparer complètement.

5.5. Conclusions

La détermination des constellations géographiques sur la base de la séparation des tendances générales, régionales, locales et temporelles permet les conclusions suivantes par rapport aux explications de Romaine et de Goeman/Van Reenen:

- 1) L'apocope du *t*-final a connu une évolution en temps réel, au moins dans les dialectes. L'apocope n'est pas un marqueur stable étant donné qu'elle a gagné du terrain en temps réel. Par conséquent la prise de position de Romaine doit être corrigée sur ce point.
- 2) Le caractère systématique des fluctuations locales est la conséquence d'une évolution *interne* des dialectes sans rapport avec la langue standard. Les tendances plus générales ne représentent ni l'influence de la langue standard (il y a une évolution inverse à l'est) ni l'influence de la distance géographique (la variable étant non pertinente) ni l'influence globale (il y a absence d'autocorrélation).
- 3) Comme la diminution de l'apocope - pour Romaine l'indice de la réintroduction du *-t* sous l'influence de la langue standard - se produit indépendamment de l'influence de la langue standard, et qu'en même temps se présente également l'augmentation de l'apocope, il n'y a aucun argument plaidoyant en faveur d'une règle de réintroduction du *t*. L'apocope est un processus uniforme, comme l'a démontré déjà Van Hout (1980).

¹⁵ Ce résultat est probablement en rapport avec la détermination de nos coordonnées, plus exacte que celle de l'analyse du tableau 2, ayant comme conséquence une diminution des erreurs.

¹⁶ Pourtant, cette influence diffuse est minimale ou absente, parce que la pondération pour le test de l'autocorrélation (à résultat négatif) rend compte d'une partie de cette influence globale.

¹⁷ C'est pourquoi un modèle sinusoïdal est peut-être plus approprié, pourtant pour tester un modèle spatial sinusoïdal il faut disposer d'au moins 100 observations (=localités) par dimension au lieu des 50 observations disponibles.

- 4) Les résultats de Goeman/Van Reenen sont compatibles avec ce que nous avons trouvé ci-dessus: l'apocope reste un processus phonologique tant naturel que sémi-naturel; la réintroduction du *t*-final ne semble possible que grâce à l'existence des alternances en-*t*- du type *knech* ~ *knechte*. Par conséquent, dans les régions où l'apocope a diminué il est question de la diminution du processus plutôt que de la réintroduction du *t*-final. En considérant la cartographie des écarts et l'absence de l'influence globale de la langue standard, nous constatons que les phénomènes de contact entre les dialectes à courte distance sont plus importants que ceux à longue distance.

L'introduction de la distinction entre les tendances globales, les tendances régionales et les tendances locales nous a permis d'obtenir des réponses précises par rapport au comportement de l'apocope dans les dialectes néerlandais.

Par leur caractère plus précis, les comparaisons des données individuelles/locales sont à préférer à celles des données groupées, comme le montre le deuxième Atlas linguistique de Dees (voir Dees et al. 1987).

