

Du français observé sous divers éclairages. Questions de corpus

La recherche sur le langage et sur les langues utilise largement des méthodes empiriques telles que celles regroupées sous l'étiquette de « linguistique de corpus » (voir Habert & al. 1977, Bilger 2000a et 2000b, Blanchet 2000). Cette démarche méthodologique revêt elle-même des formes diverses, selon le domaine, les objectifs et les conditions de la recherche. On en proposera quelques exemples, pour souligner l'intérêt, voire la nécessité en certains cas, de recueils de corpus « au vol », et leur complémentarité avec d'autres modalités de recueil¹.

Aperçu introductif

Depuis le XIXe siècle, mais plus amplement au cours du XXe siècle, le recueil et l'analyse de corpus ont fondé, informé et orienté divers domaines des recherches en linguistique : description des langues, comparaison des langues, sociolinguistique, acquisition des langues, traductologie, et bien d'autres. Et permis, dans ces domaines, des descriptions et des analyses portant sur la grammaire, le lexique et la phonologie de langues très diverses. Les méthodes de la linguistique de corpus, initialement développées sur des langues écrites ou sur la forme écrite des langues, se sont progressivement étendues aux langues parlées

¹ Ma propre expérience de recherche a largement porté sur du travail (recueil et analyse) de corpus : *Analyse de productions-réponses dans des exercices structuraux en laboratoire de langues chez des adultes germanophones* (Besançon 1966) ; *Etude lexicologique du vocabulaire des sports de ballon* (Besançon 1967) ; *Analyse d'erreurs dans les productions orales et écrites en français d'adultes anglophones* (Paris 8, 1975), *Acquisition de la référence spatiale en français par des adultes hispanophones* (Projet ESF sur l'acquisition des langues en milieu social par des adultes migrants, voir Becker & Carroll 1997, chapitre 5). D'autres travaux, plus récents ou plus ponctuels, sont évoqués dans le présent article.

et à la forme orale des langues. Plus particulièrement à partir de la fin du XIXe siècle après l'invention de l'enregistrement sonore de la parole², puis dans la seconde moitié du XXe siècle grâce aux capacités techniques du magnétophone, et plus près de nous par les moyens électroniques de stockage et de traitement de grands corpus écrits et/ou oraux³.

La place des corpus dans la recherche en description des langues et en didactique des langues, pour ce qui concerne le français et la langue parlée, a, depuis l'entreprise pionnière du Français fondamental (voir Gougenheim & *al.* 1964), connu des développements importants et divers, qu'il n'est pas possible de recenser ici. Mais ces développements même paraissent ramener à un double constat : 1° la linguistique de corpus est en grande partie une linguistique de terrain 2° la linguistique de corpus implique souvent pour le recueil une visée précise d'analyse. En bref (1°), s'il est possible de collecter ou de constituer des corpus à caractère expérimental ou semi-expérimental (par exemple des mots, des énoncés, etc. produits en situation de laboratoire), un grand nombre de corpus concernent du langage observé dans son exercice naturel⁴, et recueilli comme tel. D'autre part (2°), s'il existe des corpus ouverts, dont l'exploitation précise n'est pas prédéterminée ni préorientée (c'est le cas de Frantext), d'autres sont « ciblés », c'est-à-dire recueillis dans un objectif linguistique et/ou acquisitionnel précis, en fonction d'une dimension spécifique visée par la recherche et selon des procédures de recueil ajustées à l'objectif.

D'autre part, certains corpus sont destinés à établir des comparaisons entre des séries de données : comparaison interlinguistique (entre deux ou plusieurs langues) ou interlectale (entre variétés sociolinguistiques) ; comparaison développementale (entre divers stades d'acquisition ou

² Qui remonte à 1877 (Charles Cros et Thomas Edison, séparément).

³ Voir Habert & *al.* 1997.

⁴ Y compris en classe de langue, où les interactions d'apprentissage, dans leur diversité, sont naturelles dans le contexte de la classe : celle-ci, pas plus que les interactions qui s'y déroulent, n'est pas un montage expérimental (Il reste que la classe de langue peut être un lieu d'expérimentation didactique, ou/et d'observation de recherche). D'autre part, les « questionnaires linguistiques » (Thomas & *al.* 1971), au dispositif semi-expérimental, sont le plus souvent utilisés, sur le terrain, en complément de données plus naturelles.

d'apprentissage d'une langue, maternelle ou non maternelle, par un sujet ou un ensemble de sujets), voire interlinguistique et développementale (entre divers stades d'acquisition de langues maternelles⁵ ou non maternelles. Ce type de recherche s'accommode forcément mieux de corpus ciblés que de corpus ouverts.

Les corpus ouverts, s'ils sont suffisamment fournis, permettent assurément de traiter une grande quantité et une grande diversité de secteurs d'une langue (voir Biber & al. 1999, 1-48), y compris des structures non anticipées par la collecte, voire révélées ou suggérées a posteriori par celle-ci. Mais ils peuvent, aussi larges soient-ils⁶, ne pas fournir de données suffisantes sur certaines structures de la langue ou sur des emplois de la langue dans certaines situations de discours ou d'acquisition. Ainsi, si l'on s'intéresse, par exemple, à telle structure interrogative précise (en français, « *qu'est-ce que c'est que* + GN »), on peut en rechercher des occurrences dans des corpus ouverts sans en trouver, ou en en trouvant moins — ou différentes — que l'on en sait utilisées dans le langage ordinaire. Il n'est pas non plus possible de susciter (d'« éliciter ») la production de tel type d'énoncé. Tout au plus peut-on alors, éventuellement, fabriquer de toutes pièces des énoncés et les soumettre aux jugements de locuteurs, natifs ou non natifs, selon la visée de la recherche. Il en va de même pour constituer un corpus autour de l'expression du reproche ou du remerciement; de la référence spatiale ou de la qualification péjorative (ou, voir plus loin, sur l'insulte ou les jurons), de l'emploi de tel type de verbe ou de substantif.

La pratique du recueil « au vol » s'inscrit dans la tradition des collectes de corpus oraux et écrits au jour le jour, en dehors ou au delà des textes (littéraires ou autres) constitués. Elle est illustrée entre autres, dès le XIXe siècle, pour le français, par Darmesteter pour le lexique puis, au début

⁵ Voir Slobin, D. (1985) *The crosslinguistic study of language acquisition*. Hillsdale, Lawrence Erlbaum.

⁶ Le corpus total idéal comporterait tout ce qui a été émis par des êtres humains dotés de parole, depuis l'apparition du langage et au jour le jour ce qui est exprimé à chaque instant. Pour une seule langue (hormis celles encore parlées par seulement quelques personnes dans un espace géographique réduit), un corpus exhaustif d'une journée échapperait aux possibilités matérielles théoriques de recueil. Pour des langues à grande diffusion, il échapperait aux capacités effectives de stockage (écrit et oral) des outils électroniques existants.

du XXe siècle, par Damourette et Pichon pour la grammaire. Elle constitue une sorte de butinage réceptif systématique, qui requiert patience, attention et oreille fine. Elle consiste à noter, à chaud, « papier-crayon », les occurrences rencontrées⁷. Pour artisanale qu'elle puisse paraître à l'ère électronique, cette forme de recueil de données s'avère nécessaire — autant que passionnante — pour capter certains faits de langage et de communication. Permettant d'appréhender et de recueillir du non-prévisible, elle constitue un outil indispensable pour étudier par exemple des phénomènes de néologie, des productions spontanées d'apprenants de langue étrangère ou des formes d'interaction orale spontanée en champ libre. Et pour compléter des données recueillies par d'autres moyens.

On présentera ici quelques cas précis, pour montrer la complémentarité entre les divers types de corpus, et plus particulièrement entre les recueils préstructurés et les recueils « au vol », ainsi qu'entre les corpus ouverts et les corpus ciblés.

1. -t- comme dans flamencoter, gouroutesque, voyouterie

Ayant remarqué la productivité en français de la dérivation en *-t-* (comme dans *piano-pianoter, média-médiatique, voyou-voyouterie*) et été parfois interrogé par des professeurs ou étudiants étrangers de français sur certaines occurrences lexicales absentes de leur dictionnaire, j'ai effectué de 1977 à 1990 (Porquier 1993) un recueil 'au vol' de telles occurrences, manifestant diverses formes de dérivation incluant, à l'oral et/ou à l'écrit, un [t] ou un *t* entre une base lexicale et un suffixe⁸. Complémentairement à ce recueil, il fallait recenser dans divers dictionnaires des occurrences relevant de ce type de formation (*aléa-aléatoire, domino-dominotier*, etc.), y compris à des états antérieurs, voire anciens de

⁷ L'imprévisibilité de telles occurrences rend impraticable et/ou inutile l'enregistrement sonore. Sauf dans le cas de structures hautement prévisibles ou fréquentes dans des situations ou contextes spécifiques d'interaction ; ou, à l'opposé, inévitablement fréquentes indépendamment des contextes et des situations.

⁸ Le morphème *-t-* ([t]) sert par ailleurs à former, surtout à l'oral, du féminin, sur des substantifs et des adjectifs, de façon parfois improvisée et/ou en écart à la norme (*voyoute, blaireaute, pourrite*).

la langue. Notre propre cueillette comporte une centaine d'entrées, issues de sources diverses (conversations orales en face à face, journaux, radio, littérature, etc.) et dont certaines sont peut-être des hapax : *se métroter* » (= prendre le métro), *flamencoter* (= danser le flamenco), *gouroutesque* (= semblable à un gourou), *queneautiste* (= de Queneau). Leur diversité même permet cependant de mettre en évidence ou de préciser les mécanismes morphophonologiques et morpholexicaux sous-jacents à de telles formations, bien au delà des occurrences répertoriées par les dictionnaires consultés. De plus, les contextes d'où ils proviennent, une fois rapprochés ou regroupés, montrent l'usage stylistique qu'en font certains écrivains (tels que Céline, Cavanna et San Antonio), journalistes ou chroniqueurs, mais aussi leur traitement épilinguistique dans des improvisations lexicales lors de conversations ordinaires (« Depuis une semaine, on fait du papier. On *papiète*, quoi ! »).

Le recueil effectué (Voir Annexe 1) a montré en outre la complémentarité entre le corpus au vol (productions orales et écrites, ordinaires et/ou médiatiques) et les recensements opérables dans divers types de dictionnaires (usuels, étymologiques, dictionnaires du français d'Afrique, dictionnaires de patois ou de parlers régionaux), et avec des productions d'enfants francophones ou d'apprenants non-natifs.

2. « Je lui ai couru après, il m'a sauté dessus »

Le recueil de corpus sur la structure « je lui ai couru après », effectué de 1990 à 1997, relève du même principe. Cette structure syntaxique, signalée par les grammaires du français, et sommairement décrite en terme de transformation (montée de l'objet : 'J'ai couru après x/lui' —> 'je lui ai couru après'), est le plus souvent évoquée à propos du débat sur les statuts respectifs et comparés des prépositions et des adverbes, et sur les prépositions dites "orphelines" (« j'étais venu avec, je suis reparti sans »). S'il est difficile de considérer comme *préposition* un élément régissant une unité nominale ou pronominale qui le précède; ou absente, il est tout aussi difficile de considérer comme *adverbe* un élément régissant une unité nominale ou pronominale. D'où la nature ou le fonctionnement spécifique d'une série d'unités dans le contexte envisagé. Le recueil entrepris visait à cerner l'extension lexicale et grammaticale de la structure en question, avant d'en envisager une description plus précise (Porquier 2001a).

L'examen de corpus ouverts ne fournit là que des données partielles, dans la mesure où ces corpus sont majoritairement de sources écrites, alors que la structure en question est plus largement manifestée dans la langue parlée ordinaire⁹. C'est donc un recueil au vol, regroupant des occurrences orales et écrites dans des contextes très diversifiés, qui a été opéré, pour aboutir à un inventaire lexical (verbes, "prépositions/adverbes" et cooccurrences, voir Annexe 2). Pour opérer une analyse grammaticale précise et esquisser une analyse sémantique de l'inventaire obtenu, il était utile, pour vérifier ou compléter le corpus 'au vol', et en envisager un classement sémantico-syntaxique, de le confronter avec le dictionnaire exhaustif des verbes français de Dubois et Dubois-Charlier (1993). Il est apparu, chemin faisant, que l'italien et l'espagnol connaissent une structure comparable, et qu'une analyse comparée entre ces trois langues serait de nature à préciser le fonctionnement de ce micro-système dans chacune d'entre elles.

Ce deuxième exemple veut illustrer à la fois la nécessité en pareil cas du recueil au vol, et sa complémentarité, directe ou indirecte, avec d'autres données constituées par d'autres moyens. Il est à souligner qu'un tel recueil n'aurait probablement pas pu être effectué d'emblée sur un corpus ouvert étendu, faute de clés d'accès pour une consultation informatique, et qu'au contraire notre recensement lexico-grammatical 'au vol' fournit un levier relativement affiné pour une telle consultation.

3. Chat à la fenêtre

Parmi les nombreux travaux portant¹⁰ sur l'acquisition des moyens de la référence spatiale en langue non maternelle, les options théoriques et méthodologiques adoptées dans le projet ESF sur l'acquisition des langues en milieu social par des adultes migrants (voir Perdue 1993), étude translinguistique (portant sur une diversité de langues 'cible' et de

⁹ Même si elle ne relève pas d'un registre familier, comme le prétendent maintes grammairistes.

¹⁰ à partir de 1980. Les travaux sur l'acquisition de la référence à l'espace en langue maternelle ont été enclenchés dès les années 1930, dans le cadre de la psychologie cognitive, par Piaget. Voir Piaget, C., Inhelder, B. (1948) : *La représentation de l'espace chez l'enfant*. Paris, Presses Universitaires de France, 1948.

langues 'source') et longitudinale (itinéraire acquisitionnel sur une période de 30 mois), ont permis la constitution d'un abondant corpus spécialisé, puis son exploitation. Il s'agit d'un corpus à la fois ciblé et ouvert. Ciblé parce que constitué à partir d'une série de tâches plus ou moins libres (allant d'interactions conversationnelles authentiques à des procédures semi-expérimentales) choisies et conçues pour observer et étudier des aspects précis (structuration de l'énoncé, référence à la personne, référence au temps, référence à l'espace, etc.) du développement *ab initio* d'une compétence non-native. Ouvert car au delà des objectifs initiaux il se prête à diverses analyses permises par l'amplitude et la diversité des données recueillies.

La partie de la recherche consacrée à la référence spatiale a fourni des données substantielles et permis d'abondantes analyses (voir Becker & Carroll 1997), dont certaines apportent, dans une perspective acquisitionnelle, des éclairages utiles à la sémantique cognitive. Ainsi, pour ce qui est de la référence projective dans l'axe sagittal (du type 'devant', 'derrière', comme dans : « reste derrière/devant moi », « le ballon a roulé derrière/devant la voiture », « il y a quelqu'un devant/derrière la maison », etc.), il est apparu que le dispositif de recueil de données n'avait forcément pas pu envisager toutes les configurations spatiales potentiellement impliquées dans l'expression de telles relations, ni les entités (personnes, objets) impliquées dans la perception et la verbalisation de telles relations. Mais les questions soulevées à la fois par l'apport et par les limites des données recueillies suggèrent l'intérêt, dans d'autres cadres d'investigation, de protocoles précis, dont on présentera un aperçu.

Dans l'image ci-dessus¹¹, la position respective et relative des entités 'chat' et 'fenêtre' permet de dire « le chat est devant/derrière la fenêtre », ou autre chose, et en tous cas de verbaliser de quelque façon la perception par un locuteur de la relation spatiale entre le chat, la fenêtre et lui-même¹².

Un corpus recueilli selon ce principe ("décrire l'image") et un protocole précis apporte des données plus riches et plus fines que celles obtenues par d'autres procédures, ou recueillies au vol¹³. Aucun corpus ouvert, aucun corpus au vol ne pourraient recouvrir précisément les différents cas de figure présentés par la série d'images, elle-même conçue selon un cadre d'analyse informé au préalable par des données antérieures moins ciblées, mais aussi par un étayage théorique (Hill, Vandeloise) utile.

Les quatre images présentées en Annexe 3 amènent en français, et en d'autres langues, auprès de locuteurs natifs ou non natifs, à constater que la plupart d'entre elles se prêtent à diverses formulations, selon le point de vue (égocentrique ou hétérocentrique) adopté, et, plus précisément, que le fait de dire « devant » ou « derrière » tient, par exemple, au fait que l'on adopte le point de vue du chat (présent, lui, dans l'image), ou le point de vue externe du descripteur-observateur ; selon que la relation est perçue-appréhendée-verbalisée selon l'« imagerie déictique » en

¹¹ On trouve en Annexe 3 quatre images différentes (le protocole complet en comporte huit), sur lesquelles le lecteur peut s'exercer en au moins deux langues (sa langue maternelle et une autre). Les variables sont : 1° le fait que le chat est d'un côté (intérieur) ou de l'autre (extérieur) de la fenêtre 2° que le chat fait face ou tourne le dos 3° que l'observateur-locuteur est lui-même dans la maison ou dehors. Le principe est bien sûr que l'observateur-locuteur se trouve dans la position déterminée par l'image (l'image est ce qu'il voit, d'où il est), comme s'il en était le photographe, ou le dessinateur. Une expérience autre consisterait à faire parler le chat.

¹² La représentation adoptée, en dessin sommaire, à deux dimensions, de la configuration spatiale ne constitue pas le support idéal pour l'expérience. Mais il est bien plus difficile, expérience faite, d'imposer à un vrai chat une diversité de postures immobiles successives (le protocole complet en requiert huit) auprès d'une vraie fenêtre.

¹³ Un des résultats est que le locuteur a souvent, pour une même image, le choix entre deux formulations (« le chat est devant la fenêtre », « le chat est derrière la fenêtre ») selon qu'il "se met" à sa propre place présumée ou à la place du chat. Cette observation n'est bien sûr pas propre au français, quelle que soit par ailleurs la nationalité du chat.

‘face-à-face’ ou en ‘tandem’ (Hill 1991). Ainsi, pour l’image ci-dessus (où l’observateur-locuteur est en dehors de la maison), le chat peut, en français, être vu et dit « devant la fenêtre » ou « derrière la fenêtre » (voir également Vandeloise 1986).

On retrouve ici la complémentarité, méthodologique et heuristique, entre divers types de corpus et divers modes de recueil de données. Les méthodes expérimentales et semi-expérimentales produisent, à leur manière et à leur mesure, des corpus étroitement ciblés, à travers des protocoles de recueil ponctuels et contrôlés, là où des données ‘au vol’, au prix du naturel, ne pourraient produire que des corpus plus aléatoires dans un temps bien plus long.

4. Mots-phrases ou phrasillons, y compris des insultes

Il existe, dans toute langue, des unités lexicales qui peuvent constituer des énoncés autonomes, sans pour cela être considérées comme des réductions de phrases, comme en français : *bravo, merci, au secours, stop, tiens, ça y est*, etc. Dénommées selon les cas ‘mots-phrases’, ‘phrasillons’ (Tesnière) ou ‘locutions-énoncés’ (voir Porquier 2001b), elles n’ont pas fait, pour le français, de recensement exhaustif ni d’analyse détaillée, malgré l’intérêt linguistique et didactique qu’elles présentent. Certes, un dictionnaire d’usage les contient en principe, même s’il n’en signale pas toujours les valeurs d’emploi, mais sans en suivre à la trace les usages parfois fluctuants et sans mentionner certaines unités ou certains emplois (insultes, jurons) jugés trop grossiers. La constitution d’un corpus de telles unités suppose d’une part un tri dans les dictionnaires existants, d’autre part un relevé au vol, attentif, d’occurrences — majoritairement orales — en contexte et en situation. Ici encore, la conjonction de deux sources complémentaires s’avère pertinente et nécessaire. On aboutit ainsi à un inventaire approximatif et provisoire de cent à deux cents unités, qui se prêtent à des comparaisons interlinguistiques et à des exploitations didactiques, mais aussi à des analyses incluant les dimensions prosodiques, pertinentes du double point de vue énonciatif et pragmatique (voir Porquier 2001b, où figure également un échantillon sélectif et illustratif de telles unités, micro-corpus déjà exploité pour une description du français parlé.

Dans ce cadre s’inscrivent le recensement et l’étude des jurons, insultes

et injures¹⁴, dont un recueil systématique et exhaustif présente les mêmes limites et contraintes que ceux déjà évoqués. Au delà d'un inventaire partiel, permis par les dictionnaires et la saisie 'au vol', il reste la possibilité complémentaire d'enquête auprès de locuteurs natifs sur leur usage personnel (y compris intérieur, en soliloque, *in petto* ou *sotto voce*, sans exclure le proféré en champ libre) des jurons, insultes et injures dans des situations diverses de la vie ordinaire. Certes, les réponses d'un enquêté à un questionnaire écrit ou aux questionnements oraux d'un enquêteur (« quels jurons proférez-vous dans telle circonstance ? ») sont moins fiables que l'enregistrement sur le vif en situation réelle, mais ici encore, l'apport d'un corpus mixte s'avère précieux, voire indispensable. De plus, les distinctions entre l'insulte, le juron et l'auto-insulte, d'une part, et d'autre part quant à l'utilisation ("très souvent / souvent / rarement") et les circonstances ("devant les autres / seul(e) / dans la tête"), fournissent là une diversité de réponses, partiellement inaccessibles à l'observation externe en champ libre. On trouvera en Annexe 4 un sommaire du protocole d'enquête.

Conclusion

Ce bref survol de différents types de corpus¹⁵ et de données illustre partiellement, d'une part, la variété des procédures de recueil développées depuis quelques décennies et la diversité des domaines et des secteurs d'investigation. On constate, en même temps que cette variété, la complémentarité utile ou nécessaire, pour un même objet d'étude, de diverses modalités de recueils et de divers types de données. On remarque également qu'à cette diversité correspondent diverses postures d'observation, diverses démarches d'investigation, illustratives des

¹⁴ Qui a fait l'objet d'un travail de séminaire (projet « Jurin ») par un groupe d'étudiants de l'Université Paris-X, en 1998-1999.

¹⁵ *Corpus* : « Ensemble limité des éléments (énoncés) sur lesquels se base l'étude d'un phénomène linguistique » (*Trésor de la Langue Française*, 1978, Tome VI, p. 215) ; « Ensemble d'énoncés écrits ou enregistrés, constitué en vue de leur analyse linguistique (*Grand Dictionnaire Encyclopédique Larousse*, 1982, Tome 3, p. 2647) ; « [La grammaire descriptive d'une langue s'établit à partir d'un] ensemble d'énoncés qu'on soumet à l'analyse et qui constitue le *corpus* de la recherche ». (Dubois & al., 1994, *Dictionnaire de linguistique et des sciences du langage*, Larousse).

possibilités et/ou des contraintes de telles recherches empiriques. On aura deviné enfin qu'il y a large place, en ce domaine, pour la curiosité et le plaisir de l'observateur-chercheur.

Bibliographie

- Becker, A., Carroll, M., eds (1997) : *The acquisition of spatial relations in a second language*. Amsterdam/Philadelphia, John Benjamins.
- Biber, D., Johansson, S., Leech, G., Conrad, S., Finegan, E. (1999) : *Longman grammar of spoken and written English*. London, Longman.
- Bilger, M. (2000a) : *Corpus. Méthodologie et applications linguistiques*. Paris, Champion.
- Bilger, M., dir. (2000b) : *Linguistique sur corpus : études et réflexions*. Perpignan, Presses Universitaires de Perpignan.
- Blanchet, P. (2000) : *La linguistique de terrain*. Rennes, Presses Universitaires de Rennes.
- Calvet, L.J., Dumont, P. (1999) : *L'enquête sociolinguistique*. Paris, L'Harmattan.
- Dubois J., Dubois-Charlier F. (1993) : *Dictionnaire des verbes*. Paris, Institut Blaise Pascal, LADL.
- Gadet, F. (1989) : *Le français ordinaire*. Paris, Armand Colin.
- Gougenheim, G., Michéa, R., Rivenc, P., Sauvageot, A. (1964) : *L'élaboration du français fondamental*. Paris, Didier.
- Habert, B., Nazarenko, A., Salem, A. (1997) : *Les linguistiques de corpus*. Paris, Armand Colin.
- Hill, C. (1991) : « Recherches interlinguistiques en orientation spatiale », *Communications*, 53, 171-207.
- Nida, E. (1947) : « Field techniques in descriptive linguistics », *IJAL*, 13, 138-146.
- Perdue, ed. (1993) : *Adult second language acquisition : cross linguistic perspectives. Vol. II : The results*. Cambridge, Cambridge University Press.
- Porquier, R. (1993) : "Remarques sur la dérivation en -t- en français contemporain", *Travaux de linguistique et de philologie*, XXXI, 367-384.
- Porquier, R. (2001a) : « 'Il m'a sauté dessus', 'je lui ai couru après'. Un cas de postposition en français », *Journal of French Language Studies*, 123-134.
- Porquier, R. (2001b) : «Mots-phrases, phrasillons, locutions-énoncés », *Langue Française*, 131, 106-122.
- Revue française de linguistique appliquée* (1995) : Corpus : de leur constitution à leur exploitation, vol. 1-2.
- Samarin, W. J. (1967) : *Field linguistics*. New-York, Kolt, Rinehart & Winston.
- Thomas, J., Haudricourt, Caprile, J.P. (1971) : *Enquête et description des langues à tradition orale*. Paris, SELAF.
- Vandeloise, C. (1986) : *L'espace en français*. Paris, Seuil.

Annexe 1

Dérivation en -t- (extraits sélectifs de Porquier 1993)

acajoutier, blablater, blaireautitude, boyauterie, bureautiser, camérater, carreauté, chapiteautage, charabiater, cocacolatisé, cocoricoter, compte-renduter, créneauter, démaotisation (de Mao), drapeautique, esquimauter, farfeluteries, gagatiser, hugotesque (de Hugo), miaouteuse, noahtique (de Noah), pied-au-cu(l)ter, pinceauter, maffiaté, (se) métroter, queneautiste, reggaeter, réseauteux, sidatique, stabiloter, tangoter, tohubohuté, troupeauter, (se) véloter, vélotique, zozoterie.

Annexe 2 (Extraits de Porquier 2001a)

2.1 à côté, à droite, à gauche, après, autour, au travers, avant, (contre), dedans, (par) derrière, (au/par/en) dessous, (au/par/en)dessus, devant, entre, (pas) loin, (parmi), (tout) près.

2.2 *aboyer, aller, appuyer, arriver, atterrir, avancer, basculer, baver, beugler, bondir, bourrer, cavalier, chier, cogner, coller, couler, courir, cracher, crier, danser, débarquer, débouler, décharger, déferler, dégobiller, dégouliner, dégringoler, dégueuler, descendre, dévaler, éjaculer, faire, foncer, fondre(2), gicler, glisser, goutter, grimper, grogner, gueuler, hurler, jouir, lancer, marcher, monter, neiger, passer, peser, pisser, pleurer, pleuvoir, plonger, pousser, rebondir, remonter, rentrer, rester, retomber, rouler, saigner, sauter, siffler, shooter, souffler, taper, tirer(2), tomber, tourner, trotter, uriner, venir.*

2.3 (Extraits sélectifs)

- *Et à chaque fois ça me retombe dessus, c'est pas à lui que ça retombe dessus, tiens* [oral, 02-04-95, dans un autobus]
- *Tous les matins je je j'ai des femmes avec enfants dans la voiture qui euh me rouleraient dessus sur les passages cloutés* [oral, corpus Flament 12/21/94]
- *Il lui a grimpé dessus sous la table de la cuisine* [oral, femme 70 ans, outrée, février 97 à propos de chiens]
- *Arrêtez de me neiger dessus* [oral, jeune homme s'adressant en riant aux flocons de neige, 27-12-96]
- *Il nous a passé entre* [football, interview, radio, s.d.]
- *"Ils m'ont chié dessus avec cette histoire de viol"* [Depardieu, *L'Echo des Savanes*, février 97, cité, à propos d'un article de Times Magazine]
- *Ce que nous ne comprenons ni Mary ni moi, c'est leur manie de se grimper dessus, surtout les petites bête [...]. Il n'y a pas que les petites bêtes qui se*

grimpent dessus, les moyennes aussi.[R. Queneau, *Les œuvres complètes de Sally Mara. Journal intime*. Paris, Gallimard, NRF, 1962]

- *"Après j'ai les dossiers qui vont me débarquer dessus"* [22-2-97, oral, rue]
- *Mais c'est vous qui me passez devant, restez à votre tour* [oral, altercation, queue dans magasin, 28-11-96]
- *A l'entraînement, il lui courait après, enfin, au deux sens du terme* [Club sportif, oral, s.d.]
- *Tu viens vers moi, tu me tournes autour et tu repars même sens* [Leçon de danse, oral, s.d.]
- *Le voilà qui lui recourt derrière* [récit oral, poursuite, 11-12-95]
- *Avec une grande robe style viennois qui volait qui lui flottait autour* [récit oral de film, 11-12-95]
- *Je m'arrête de fumer et voilà dix millions qui me tombent dessus* [*Charlie Hebdo*, 4-9-97]
- *Là, il arrêta pas de me coller après, pendant des semaines* [récit oral d'aventures sentimentales, métro, 14-3-96]
- *Et vous mes mains ne tremblent plus, souvenez-vous quand je vous pleurais dessus* [Chanson *Mathilde*, Jacques Brel]

Annexe 3

(Les quatre dessins ci-dessous font partie de l'ensemble de huit dessins minimalement requis pour le protocole "Chat à la fenêtre")

Annexe 4 (Jurons, insultes, injures)

1. En voiture

- lors d'un accrochage ou d'un accident
- dans les embouteillages
- vis-à-vis de piétons ou cyclistes

2. A la maison, en famille

- lors d'une dispute
- quand on casse ou renverse quelque chose
- vis-à-vis d'un animal

3. Au travail

- quand on a raté ce qu'on fait
- quand on est dérangé au mauvais moment
- lors d'un conflit avec un collègue

4. Divers

- au téléphone, en raccrochant énervé
- en réaction à une bousculade ou une agression dans un lieu public
- quand on a perdu ou oublié quelque chose

5. Autres.