

Les *corpus dynamiques* et GlossaNet. Extension de la couverture lexicale des dictionnaires électroniques anglais

Cédric Fairon, Blandine Courtois

Laboratoire d' automatique documentaire et linguistique (LADL)
Université Paris 7 - 2 Place Jussieu (case 7031) - 75251 Paris CEDEX 05

Abstract

Electronic dictionaries play a key role in the automatic language processing systems developed at the LADL. These dictionaries are in fact wide coverage databases which provide morphological, syntactic and semantic information for each lexical entry. Maintaining and extending these dictionaries is a considerable task. In this paper, we present an automated system that simplifies this maintenance. This system, based on the parsing of *dynamic corpora*, has been used since September 1999 to update the English electronic dictionaries of the LADL.

Résumé

Les dictionnaires électroniques sont un constituant central des systèmes de traitement automatique du langage développés au LADL. Ces dictionnaires sont des bases de données à très large couverture fournissant pour chaque entrée lexicale une série d' informations morphologiques, syntaxiques et sémantiques. L' entretien de tels dictionnaires et leur extension progressive représente un travail considérable. Nous décrivons dans cet article un système d' exploitation d'*corpus dynamique*, testé depuis septembre 1999 et destiné à faciliter l' actualisation des dictionnaires électroniques de langue anglaise du LADL.

Keywords: lexicography, electronic dictionaries, press corpora, dynamic corpora, maintenance of dictionaries, INTEX, GlossaNet, CorpusWeb.

1 Introduction

Les dictionnaires électroniques sont un constituant central des systèmes de traitement automatique du langage développés au LADL. Ces dictionnaires sont des bases de données à très large couverture lexicale fournissant pour chaque entrée une série détaillée d' informations morphologiques, syntaxiques et dans certains cas sémantiques (B. Courtois, 1990, 1999). Dans le cadre du réseau de laboratoires RELEX¹, des dictionnaires respectant les mêmes conventions ont été construits pour le français, l' espagnol, l' italien et le portugais. D' autres langues sont actuellement en cours de développement : le grec, le norvégien et le slovaque.

La taille et la précision de ces dictionnaires font que leur entretien et leur extension progressive demandent un travail considérable. Pour faciliter cette tâche, nous avons mis au point un système qui propose à intervalles réguliers de nouvelles entrées pour le dictionnaire en se basant sur l' exploration d' *corpus dynamique*.

¹ Les laboratoires européens membres du réseau RELEX travaillent chacun sur leur propre langue en utilisant un système de description commun : le système de représentation des dictionnaires DELA. Voir au sujet de RELEX : <http://www.ladl.jussieu.fr/Relex/RELEX.html>.

2 Corpus Dynamiques

Quand on parle de corpus, on fait généralement référence à une collection de textes qui a été établie de manière définitive sur la base de critères choisis en fonction des applications envisagées pour le corpus. Cet ensemble de textes n' a pas vocation d' évoluer au cours du temps, c' est une base de données statique utilisée dans la recherche en linguistique pour fonder les études sur une réalité tangible et expérimentable (D. Biber, 1998).

Il est cependant possible d' envisager les corpus d' un point de ~~vue~~ non-statique. Nous empruntons à A. Renouf (1994) la notion *de corpus dynamique* pour évoquer « un flux de données textuelles sous forme électronique, tel que les textes de journaux, par exemple ». A. Renouf a développé ce concept dans le cadre du projet de recherche AVIATOR dont le but était de mettre au point un système capable d' étudier l' évolution d' un corpus au cours du temps (nouveaux mots, nouveaux sens, nouvelles combinaison de mots, évolution des fréquences, etc.). L' évolution du journal *Times* a été étudiée grâce à ce système entre 1989 et 1993.

Le système que nous avons utilisé, GlossaNet (C. Fairon, à paraître en 2000), exploite le concept de *corpus dynamique*, mais d' une manière qui lui est propre. En effet, la gestion du corpus est entièrement assurée par le logiciel qui télécharge quotidiennement sur Internet l' édition du jour du journal étudié, et archive les anciennes éditions. Le flux de données est donc obtenu par un système automatique de type *pull*. Deux caractéristiques supplémentaires permettent de définir notre logiciel. Premièrement, il s' agit d' un système autonome : une fois paramétré et initié, il réitère ses tâches de téléchargement et d' analyse à intervalles réguliers sans intervention extérieure, puis les résultats des traitements effectués sont adressés par courrier électronique à un *veilleur*. Deuxièmement, il intègre les programmes du logiciel INTEX ² (M. Silberztein, 1993) ce qui permet d' effectuer des analyses relativement fines des corpus.

3 GlossaNet

L' utilisation que nous présentons ici de GlossaNet n' est qu' une application du système parmi d' autres. A la base GlossaNet a été développé pour mettre sous surveillance un site Web dont le contenu varie au cours du temps. Il permet de réaliser des opérations de veille, mais aussi faire de l' extraction d' information et de rechercher des structures lexicales ou syntaxiques. C' est cette dernière fonctionnalité que nous allons décrire.

3.1 Processus général

Le schéma présenté ci-dessous illustre le processus complet d' analyse d' un site Web à l' aide de GlossaNet. Tout d' abord, un module spécialisé CorpusWeb ³, télécharge le site Web et le convertit à l' aide d' un filtre en un corpus de texte simple, prêt à être analysé par un logiciel de traitement de corpus. Ensuite, l' analyse du corpus est réalisée à l' aide des programmes INTEX : application de dictionnaires et de grammaires locales, recherche par expressions régulières ou graphes (M. Silberztein, 1999), etc. Des programmes indépendants peuvent être ajoutés par l' utilisateur. Par

² INTEX est un analyseur de corpus développé par Max Silberztein au LADL. Ce logiciel intègre des descriptions linguistiques à large couverture (dictionnaires électroniques et grammaires locales) et permet d' effectuer des recherches de structures lexicales et syntaxiques à l' aide d' expressions régulières et de graphes à états finis.

³ CorpusWeb est également diffusé indépendamment de GlossaNet. Une version Beta du logiciel est accessible gratuitement sur <http://glossa.ladl.jussieu.fr>. Ce logiciel peut télécharger n' importe quel site Web et le transformer en corpus. Il peut être lancé en ligne de commande, ce qui permet de l' intégrer dans d' autres programmes. Un manuel technique accessible sur le même site détaille les particularités de ce logiciel.

exemple, en ajoutant un programme de conversion d' image (conversion d' un fichier BMP en JPG) nous affichons au jour le jour, dans une page Web, les graphiques résultant de l' analyse statistique par INTEX de l' édition du journal *Le Monde* (fréquence des mots composés) ⁴. Enfin, les résultats sont soit envoyés par courrier électronique (grâce au module *sendmail*), soit enregistrés dans une base de données (grâce au module *store*). Cette dernière possibilité peut, par exemple, servir à alimenter une base de connaissance.

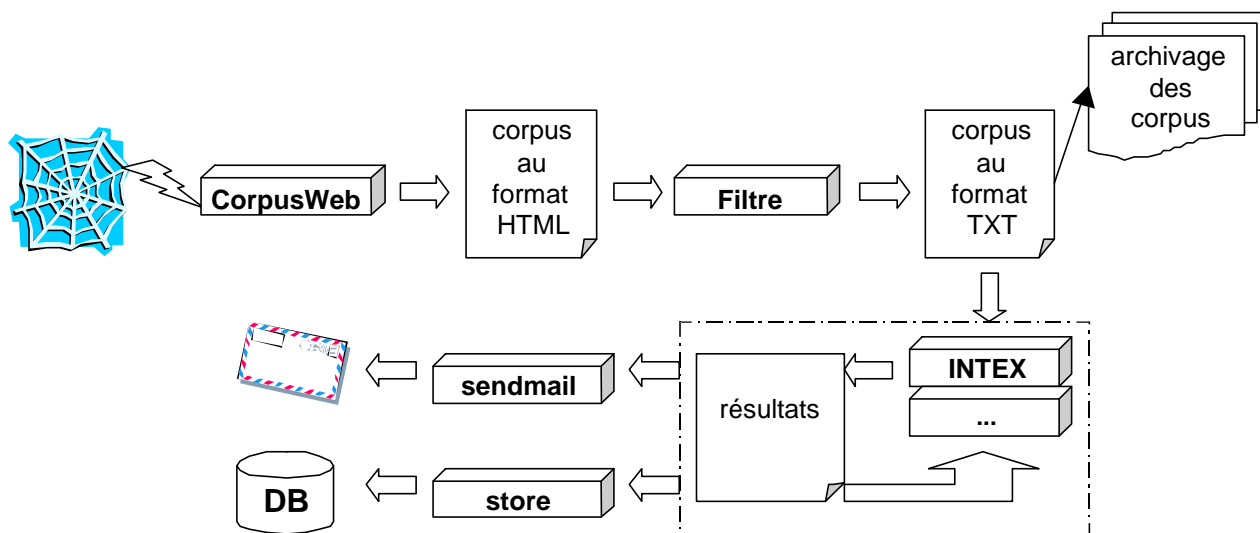


Figure 1 : Processus général de GlossaNet

3.2 Enrichissement des dictionnaires électroniques

Depuis septembre 1999, GlossaNet est utilisé au LADL pour enrichir les dictionnaires électroniques DELA de l' anglais. L'opération se déroule en deux temps.

La première phase est prise en charge par GlossaNet. Le logiciel télécharge quotidiennement les éditions électroniques du *New York Times* et du *Washington Post* et analyse les nouveaux corpus constitués, à l' aide des programmes du système INTEX. Sur chacun des corpus sont successivement appliquées quatre ressources lexicales : un dictionnaire filtre, le dictionnaire DELAF anglais, un graphe traitant les URL (ces adresses électroniques sont fréquentes dans les éditions en ligne des journaux) et un graphe reconnaissant les noms propres ⁵. Les mots reconnus sont étiquetés et enregistrés dans un fichier et les mots non reconnus sont stockés dans un autre. Ensuite, les mots non reconnus sont recherchés dans le corpus. Enfin, une concordance présentant leurs occurrences est construite avant d' être envoyée par courrier électronique à un lexicographe.

La deuxième phase est prise en charge par le lexicographe. Périodiquement, celui-ci rassemble les concordances relevées dans un journal particulier, puis étudie en détails les résultats de façon à sélectionner les candidats valables qui seront ajoutés aux dictionnaires électroniques. Cette

⁴ Ces graphiques peuvent être visualisés sur <http://glossa.ladl.jussieu.fr/stat/demostat.html>.

⁵ Ce graphe est appliqué sur le texte après le passage des dictionnaires. Il formule l' hypothèse selon laquelle les mots qui n' ont été reconnus ni par le dictionnaire DELAF ni par le dictionnaire filtre sont des noms propres s'ils commencent par une majuscule.

seconde phase n' est pas automatisable. On s' en rend vite compte en observant la feuille de concordance ci-annexée qui offre un échantillon de matériaux extraits du *Washington Post*, sur lesquels le lexicographe travaille (pour des raisons de lisibilité, les mots « inconnus » sont recopiés en tête de ligne par un programme de formatage).

Les mots sélectionnés par le lexicographe sont codés, puis enregistrés dans le dictionnaire filtre de sorte qu' aux prochains lancements de GlossaNet, ces mots ne soient plus considérés comme inconnus.

4 Étude des concordances

Dans tout dictionnaire, qu'il soit électronique ou sur papier, l' ajout d' un mot ne saurait se faire sans un examen critique préalable. Cet examen est réalisé sur le fichier de concordance fourni par GlossaNet, qui présente les mots « inconnus » avec le contexte gauche et droit (cf. Annexe).

Pratiquement le contexte est indispensable dans plusieurs cas. Premièrement, une forme graphique « inconnue » peut être une erreur typographique ou un mot mal orthographié, et de telles erreurs sont plus facilement repérables à l'intérieur d'une suite de mots que présentées isolément. Deuxièmement, plus important, un mot ne prend son sens qu'à l'intérieur d'une phrase et son identification grammaticale ne peut être déterminée sans contexte. Certes, quelques formes isolées, par exemple les noms pluriels en *-ings* ou les adverbes en *-ably*, sont catégorisables sans ambiguïté. Par contre, d'autres formes comme celles de terminaison *-ing* peuvent appartenir à différentes catégories (participe présent, substantif, adjectif), si bien que le contexte devient indispensable pour les appréhender. Troisièmement, les parties simples de mots composés deviennent parfois incompréhensibles quand elles sont détachées du composé dans lequel elles apparaissent. De même, les abréviations hors de leur contexte s' interprètent difficilement. Pour toutes ces raisons, il est important de présenter les occurrences de mots « inconnus » avec le contexte dans lequel ils ont été trouvés.

4.1 Examen préliminaire

Avant de commencer cet examen, rappelons que l'objectif premier est d'enrichir la couverture lexicale du DELAF, dictionnaire des formes simples de l'anglais. Donc, les formes relevées sont uniquement des mots de graphie simple, c'est-à-dire sans séparateurs tels que le trait d'union, l'apostrophe ou le point. Cela a pour conséquence que les mots composés sont découpés en leurs constituants simples.

La présence du contexte permet d'éliminer dès le départ les concordances qui proviennent de textes rédigés dans une langue étrangère. Par exemple, il arrive que le *Washington Post* publie des textes en espagnol, d' où le mot *propio* en fin de la page d' exemple. En outre, sont éliminées quelques occurrences qui semblent être des erreurs typographiques : *provacy* (*privacy* ?), *playoffs* (*playoffs* ?), *predeliction* (*predilection* ?).

Après suppression de ces concordances « parasites », celles qui restent peuvent être réparties d'après leur aspect formel en :

- formes simples, telles que *personas*, *phentermine*, *pianissimos*, *playmaking*, *prosecutably*, etc.
- formes entrant dans la constitution de composés, comme *phen* (constituant de *fen-phen*, *prowed* (portion de *knife-prowed*)).

Un ordre de grandeur du nombre de mots à examiner après ce premier balayage et le « nettoyage » des concordances, peut être indiqué. A partir de 18 numéros du *Washington Post* téléchargés, il est resté 1700 lignes de concordances comprenant 610 mots simples et 210 formes constituantes de composés.

Analysons maintenant dans le détail les formes de la page donnée en exemple.

4.2 Formes simples

L' échantillon proposé illustre la diversité des mots et formes trouvés :

- mots de formation régulière et de diverses catégories grammaticales : *proliferator* (nom), *prosecutably* (adverbe), *psychographic* (adjectif) ;
- formes plurielles : *personas*, *pianissimos* (le DELAF possédait déjà les entrées *persona* et *pianissimo*, avec *personae* et *pianissimi* comme formes plurielles) ;
- termes scientifiques, relevant de domaines spécialisés, par exemple *phentermine* ;
- adjectifs construits sur des noms suivis du suffixe *-ed* : *pickpocketed*, *poneytailed* ;
- noms composés soudés : *planemaker*, *postseasons*, *profitmaking*, *popup* ;
- termes français écrits sans diacritique : *premiered*, *protege*, *pere* ;
- substantifs en *-ing* : *polygraphing*, *profitmaking* ;
- formes adjectivales au comparatif ou superlatif : *prettier*, *prettiest* (le DELAF contenait déjà l'adjectif *pretty*, mais avec un code insuffisant pour obtenir le comparatif et le superlatif) ;
- mots construits à l'aide de préfixes soudés: *postseasons*, *prefight*, *premodernist* ;
- abréviations : *pol*, *pols*, *prez*, *pp*, *ppg*.

Enfin, certaines graphies sont la transcription du langage parlé : *poppin'* and *lockin'*. Dans ces formes, le *g* final de la terminaison *-ing* est remplacé par une apostrophe. Signalons aussi le procédé qui consiste à écrire *poooooow* avec la voyelle répétée pour renforcer l'intonation. Ces dernières formes ne sont pas retenues comme entrées de dictionnaire.

La soudure des mots composés binaires est un mécanisme important à noter, car c' est une source extrêmement fréquente de non-reconnaissance d'un mot, bien que ce dernier puisse se trouver dans le dictionnaire électronique DELACF contenant mots et formes composés de l'anglais. En attendant de trouver une solution à ce problème, il est nécessaire d'inclure les composés soudés dans le dictionnaire des mots simples.

4.3 Formes issues de mots composés

Les concordances font ressortir des constituants de mots composés. Grâce au contexte, les composés eux-mêmes sont mis en lumière. Ils apparaissent soulignés sur la feuille-exemple jointe.

Parmi les composés trouvés sur l'ensemble du *Washington Post* traité, on distingue les types suivants :

- mots composés usuels : *sure-footedly*, *do-gooders*, *left-behinders*, *fast-foodery*, *touchy-feely*, *topsy-turvy*, *boogie-woogie*, *cha-cha*, *hi-fi*, *kung fu*, *left-hander*, *right-hander*, *glad-hander* ;

- particules et mots grammaticaux au pluriel : *sing-alongs, ifs and maybes, have-nots, twenty-somethings, so-and-sos, thank-yous, add-ons, 12-and-unders* ;
- mots latins du vocabulaire scientifique : *pfisteria piscicida, archaeoraptor liaoningensis, aloe vera* (substance), *beipiaosaurus inexpectus, aedes albopictus* ;
- mots français adoptés en anglais : *fait accompli, nuee ardente, cause célèbre, chaise longue, joie de vivre, je ne sais quoi, déjà vu, laissez-faire, papier mache, nouveau riche, du jour* ;
- locutions étrangères notamment d'origine latine *in absentia, sotto voce, magna cum laude, reductio ad absurdum, terra firma, per favore, al dente* ;
- expressions mixtes constituées de mots anglais et étrangers : *pere and fils, Culex pipiens mosquito*.

GlossaNet se relève donc aussi intéressant pour la collecte de mots composés que pour celle des mots simples, même si des adaptations spécifiques aux mots composés devraient être envisagées pour un meilleur rendement. Il fournit en outre des attestations de composés au pluriel, comme *alma maters*, dont la forme plurielle n' est généralement pas donnée dans les dictionnaires.

5 Conclusion

Le dictionnaire DELAF anglais qui a été utilisé comportait à la base près de 300.000 entrées. L' expérience menée sur trois semaines de texte du *Washington Post* a permis d' ajouter plus de 800 formes nouvelles.

Des comparaisons de corpus à plus large échelle ont été réalisées par ailleurs pour le français (M. Silberstein, 1995 ; C. Fairon 1998). Elles nous laissent penser que ce rendement devrait se maintenir pendant plusieurs années. Par exemple, la comparaison des hapax apparaissant dans le quotidien belge *Le Soir* des années 1994 (*ls94*) et 1995 (*ls95*) donne les chiffres suivants : sur les 113.487 hapax relevés dans *ls94* et les 113.693 relevés dans *ls95*, seuls 17.206 hapax sont identiques⁶. Ce qui peut se représenter de la manière suivante :



Outre le rendement intéressant en nombre de mots nouveaux et l' assistance offerte aux lexicographes, le système GlossaNet offre la possibilité de confronter constamment le contenu des dictionnaires existants avec le vocabulaire rencontré dans les journaux quotidiens. Cela permet non seulement de suivre l' évolution du langage réel, tel qu' il est pratiqué au jour le jour, mais aussi de vérifier le bon fonctionnement du dictionnaire pour les entrées existantes. Cet outil offre donc une aide précieuse pour le développement des dictionnaires électroniques.

⁶ Dans ces listes d' hapax apparaissent bien sûr un certain nombre de noms propres ou de mots mal orthographiés qui ne sont pas pertinents pour notre étude, mais aussi un grand nombre de mots appartenant au vocabulaire général ou à un vocabulaire de spécialité.

Références

- Courtois B. (1990). Dictionnaires électroniques du français. In Courtois B. and Silberztein M. editors, *Dictionnaire électronique du français*. Langue française N°87: 11-22. Larousse. Paris.
- Courtois B. (1999). *Dictionnaires électroniques anglais DELAS et DELAF*. Rapport technique LADL. Université Paris 7.
- Fairon C. (1998). La presse belge au crible de l' analyse automatique. Application du système INTEX. In Klein J., Lamiroy B. and Pierret J.-M. editors, *Théorie Linguistiques et applications informatiques. Actes du 16^e colloque européen sur la grammaire et le lexique comparés (24-27 septembre 1997)*. Cahiers de l' Institut de linguistique de Louvain-Louvain-la-Neuve.
- Fairon C. (à paraître en 2000). Le système GlossaNet. In Fairon C. editor, *Le système INTEX, applications et développements. Actes des journées INTEX (1996 et 1999)*. John Benjamins. Amsterdam/Philadelphie.
- Renouf A. (1994). Corpora and Historical Dictionaries. In Lancashire I. and Russon Wooldridge T. editors, *Early Dictionary Databases*. Center for Computing in the Humanitie. University of Toronto. Pages 219-235.
- Silberztein M. (1993). *Dictionnaires électroniques et analyse automatique de textes. Le système INTEX*. Masson. Paris.
- Silberztein M. (1998). Transducteurs pour le traitement automatique des textes. In Lamiroy B. editor, *Le Lexique-grammaire*. Travaux de Linguistique N°37: 127-142. Duculot. Bruxelles.
- Silberztein M. (1995). Dictionnaires électroniques et comptage de mots. In Actes du colloque JADT' 95 : Statistical Analysis of Textual Data, Rome.
- Biber D., Conrad S., Reppen R. (1998). *Corpus Linguistics. Investigating Language Structure and Use*. Cambridge Approaches to Linguistics. Cambridge University Press.

Annexe : Concordances extraites du Washington Post (septembre et octobre 99)

pere	/ this is an oblique knock at the Cookes,	<pere>	and fils, echoed by many in the n/24
perservered/f	the season in Rockingham, N.C. Martin	<perservered>	as rookie Tony Stewart out/27
personals	/-school professor, attracted to her by a	<personals>	ad her helpful sons placed i/29
personas	/bber Duck and his CB radio. Fans of both	<personas>	packed the Birchmere on Tuesd/29
=	/onal rollout, the road-testing of public	<personas>	and spin machines before a fe/29
=	/ in New York to help me manage my online	<personas>	. Microsoft Corp. unveiled its/14
=	/reme Court justices, they were men whose	<personas>	were as oversized as their am/24
pfiesteria	/dustry has yet to fully recover from the	<pfiesteria>	outbreak two years ago and /30
phen	/ 1996: Redux approved by FDA. 1997: Fen-	<phen>	is taken off the market after stu/12
=	/ound the clock to keep up with the Fen-	<phen>	demand. Atkins reissues low-carb/12
phentermine	/s show a combination of fenfluramine and	<phentermine>	produces weight loss; fact/12
pianissimos	/its proper length and dynamic range. Her ++	<pianissimos>	are thrilling, and he/25a
=	/its proper length and dynamic range. Her	<pianissimos>	are thrilling, and her rai/25b
piching	/Thompson first-base coach, Dave Righetti	<piching>	coach, Carlos Alfonso administ/14
pickoff	/ming win over Oakton. Simpson's 45-yard	<pickoff>	return down the left sideline /16
=	/f completions, either." Simpson's first	<pickoff>	could not have come at a bette/16
pickpocketed/	, 2600 block, 6 p.m. Oct. 16. A male was	<pickpocketed>	of a wallet containing ca/28
pipiens	/s is primarily transmitted by the Culex	<pipiens>	mosquito that is active from d/28
=	/The virus is spread mostly by the Culex	<pipiens>	mosquito, which weighs about 1/29
=	/cials began capturing thousands of Culex	<pipiens>	several weeks ago in Baltimore/29
piscicida	/mption of seafood for fear of Pfiesteria	<piscicida>	, while nearly two-thirds of /30
pivotman	/-play grounder, Guillen threw too low to	<pivotman>	Lockhart who double-clutched /25b

JADT 2000 : 5^{es} Journées Internationales d'Analyse Statistique des Données Textuelles

= /lay grounder, Guillen threw too low to ?? <pivotman> Lockhart who double-clutched /25a
 planemaker /s long-range routes, giving the European <planemaker> its fifth victory over Boei/27
 platooning /ent offensive woes could result in more <platooning> of the quarterbacks during /30
 playmaking / Void provide the Hornets with speed and <playmaking> ability. Burns returned two/29
 playoffs /lay Richard Montgomery today in 3A West ? <playoffs> . (c) Copyright 1999 The Washi/28
 plein /-Sunday; closes Sunday. Impressionistic " <plein> air" oil paintings by Procter an/30
 pol /of the gun lobby, as a fat-cat moneybags <pol> , as the black-hatted sheriff blowi/12
 pols /r commission has been a place to appoint <pol> and bigwigs to produce an occasio/24
 polygraphing/several other individuals as a result of <polygraphing> . One person involved in /16
 ponied / producing companies--CBS and Paramount-- <ponied> up the dough to have "I Am the /24
 = /estate firms and other service companies <ponied> up \$3,000 apiece to be listed a/28
 ponytailed /urns 21. On a back courtroom bench, the <ponytailed> victim sat silently in the /24
 pooching / shorter off the tee, you'd think I was ! <pooching> the ball. Wilbon is a big hi/12
 poooooow /me about ego. I was probably going like ! <poooooow> . I bothered some persons. I fo/16
 poppin /ance extravaganza. The break-dancers are <poppin> ' and lockin', the tappers are t/29
 popup /y was throwing great. We got beat with a <popup> to right field, a Yankee Stadium/27
 poseur / to law-enforcement officials, the FBI <poseur> and hots eattle continued to cha/23
 postseasons /-back sweeps say it. Those 11-2 and 11-1 <postseasons> say it. The degree to whic/28
 postsunset /obbying from commissioners to stay alive <postsunset> . Commission opponents say /24
 powerlifting/hes and assistant coaches are needed for <powerlifting> , soccer, volleyball and b/14
 pp / By Andrew Delbanco Harvard Univ. 143 <pp> . \$19.95 BUILDING A BRIDGE TO THE /25b
 = /OUTCAST By Jose Latour Akashic. 217 <pp> . Paperback, \$13.95 By Richard Lipe/12
 = / Future By Neil Postman Knopf. 2-13 ** <pp> . \$24 From the moment the fir /25a
 ppg / The Greenbelt school is averaging 38.8** <ppg> . Next: Northwestern, Sat., 1:30 /25a
 = / The Greenbelt school is averaging 38.8 <ppg> . Next: Northwestern, Sat., 1:30 /25b
 predeliction/despite his talkative manner and his --- <predeliction> for hugging strangers--few/28
 preflight /ng through an interpreter at Wednesday's <preflight> news conference, declined any/24
 prefs /mend you back up important files such as <prefs> .js before editing them. We are /28
 = /ape\Users\your name) and open the file " <prefs> .js". You will see references to /28
 prego /l works such as Willaert's "Un giorno mi <prego> una vedovella," sufficiently /16
 = / steaming pasta and crying, "Vino rosso, <prego> !" Seriously, sometimes the obvio/25b
 premiered /(performed last night by the pianist who <premiered> it, Emanuel Ax) that uses a /24
 = /f the Big Apple Circus's Bello & Friends <premiered> in Dulles last week, but it'/30
 premodernist/movie is a postmodernist construction by <premodernist> comedy minds (Alan Zweibe/15
 prenups /ions of our day. No mention of money--or <prenups> either. Highlights of the eve/16
 prestart /ttee's refusal to give him time to fix a <prestart> rudder problem. He lost the r/28
 = /sman, Read, who bore off at speed during <prestart> maneuvers Saturday and T-bone/27
 = / the stern of AmericaOne at speed during <prestart> maneuvers, spinning AmericaOn/29
 prettier /reference, there will always be somebody <prettier> and she'll never value hersel/15
 prettiest /ve feet long. Now, the moray is not the <prettiest> of God's creatures, equipped/27
 = /you'd spend the time driving some of the <prettiest> country roads your radials a/27
 = /erms of pure sound, it wasn't always the <prettiest> ensemble or the most virtuos/25b
 prez / pursue a high-priced call girl, and the <prez> and his VP clash publicly (Channe/29
 profitmaking/mer president of the Sundance Group, the <profitmaking> part of Robert Redford's /27
 proliferator/. North Korea, arguably, is the largest <proliferator> of missiles and enabling /14
 pronouced /n person. He had a gunshot wound. He was <pronouced> dead at 4:41 a.m. ROBBERIES/28
 propio /dos y perdido 42. Este año batio ademas su ---<propio> record con 11 victorias cons/16
 propter / was a classic case of the post hoc ergo <propter> hoc (after this, therefore bec /12
 prosecutably/o see whether the secretary had made any <prosecutably> false statements. Ms. Bru/15
 protege /pularity. Habibie is a longtime Suharto <protege> and cabinet member who used to /16
 = /t-down of faux rock stars (former Reznor <protege> Marilyn Manson and perennial pe/29
 provacy /r. But lawmakers were so divided on what --- <provacy> provisions should be enact /16
 prowed /The contestants ride candy-flaked, knife- <prowed> boats bedecked with sponsors' de/14
 psychographicas UPN's faux wrestling show, while the <psychographic> on "Action" is much tonie/15
 = /y it's because "Family Guy" has the same <psychographic> as UPN's faux wrestling s/15
 pugil /r a Marine died in a training fight with <pugil> sticks. And in 1989, a Marine /16