PRIMERS APUNTS SOBRE ELS RESULTATS DE LA IMMERSIO EN ALUMNES CASTELLANOPARLANTS.[x]

JUAN ANTONIO CORDERO ALONSO

[x] Per a la realització d'aquest treball hem comptat amb el suport d'una borsa d'estudi donada per Abast i el Museu de L'Hospitalet.

INDEX TEMATIC.

1. INTRODUCCIO. 1
    1.1. Punt de partida. 2
    1.2. Importància didàctica del lèxic. 9
    1.3. El lèxic com aspecte relegat. 11
2. LITERATURA PREVIA. 13
    2.1. Aportacions globals i metodològiques. 13
    2.2. Aportacions lèxiques específiques. 15
3. DISSENY. 20
    3.1. Anàlisi del problema. 21
    3.2. Objectius. 22
4. METODOLOGIA. 22
    4.1. La mostra. 22
    4.2. Procediment. 23
    4.3. Segmentació d'unitats gràfiques 25
    4.4. Processament de les dades. 26
5. RESULTATS. 27
    5.1. Vocabulari. Paraules i vocables. 27
        5.1.1. Comparació dels corpus. 27
        5.1.2. El vocables obtinguts. 28
        5.1.3. Redundància lèxica. 30
    5.2. Vocables menys i més actius. 32
        5.2.1. De Freqüència 1. 32
        5.2.2. Els 50 de més freqüència. 33
    5.3. Longitud mitja de la paraula. 35
    5.4. Solapaments lexicals. 38
    5.5. Polisèmia, ús i longitud de paraula. 41
    5.6. Interferències lingüístiques. 42
    5.7. Conclusions. 44
6. DISCUSSIO GENERAL. 47
7. BIBLIOGRAFIA. 50

INDEX D'ILUSTRACIONS POR ORDRE D'APARICIÓ.

Gràf 1. Paraules emeses 27
Gràf 2. Rangs i text cobert 28
Gràf 3. Vocables produïts 29
Gràf 4. Redundància lèxica 31
Gràf 5. Vocables de freqüència 1 32
Gràf 6. Longitud mitja dels vocables 36
Gràf 7. Longitud i vocables 37
Gràf 8. Vocables específics 39
Gràf 9. Interferències: paraules 42
Gràf10. Interferències: ocurrències. 42

INDEX DE TAULES PER ORDRE D'APARICIÓ.

Taula I. Redundància lèxica 30
Taula II. Vocables més freqüents 33
Taula III. Longitud, polisemia i freqüència 41

1. INTRODUCCIO.

Des de sempre l'home estat preocupat per conèixer el funcionament del llenguatge com mitjà de comunicació i l'intercanvi d'idees, pensaments, sentiments, etc. Unes vegades es van prioritzar enfocaments descriptius i analítics propis de la concepció del llenguatge com a producte, altres ha tingut més pes anàlisis de tipus psicològics i competencials propers a la concepció del llenguatge com a un procés.

De fet el tema del llenguatge, abordat tant des d'una perspectiva estrictament lingüística com a psicològica, ha produït una gran quantitat de treballs i investigacions. A més, als últims anys, i degut a diverses raons de tipus sociopolític i històric en les que no entrarem, s'ha treballat molt sobre la problemàtica referent a la existència de dos codis lingüístics per un mateix parlant, es a dir el tema del bilingüisme. S'ha tractar d'aclarir de quina manera es millor portar a terme els aprenentatges de totes dues llengües, de quina manera es pot arribar a aconseguir una competència lingüística equiparable entre la llengua materna i la segona llengua, etc, etc.

De tots aquests estudis, n'han sortir més preguntes que respostes. Sense dubte el tema es molt complex. La gran quantitat de variables intervinients (memòria, motivació, capacitat de processament, etc), les diferències entre llengües (tant pel que fa a la estructura com al prestigi o l'equilibri d'una respecte a l'altre), el nivell cultural del bilingüe, son parcialment responsables d'aquesta gamma de resultats moltes vegades contradictoris. També ho són els interessos, els prejudicis ideològics, les perspectives, les metodologies emprades i el objectius diferents dels que partien les investigacions.

En resumen, sembla que no temin una base de coneixements i interpretacions dels mateixos, suficientement consolidada i verificada, es a dir una marc teòric respectat por tothom.

Nosaltres pretenem fer un estudi lingüístic en un interval de temps on es consoliden els aprenentatges lingüístics bàsics. Aquest període es capaç de reflectir la evolució de les variables lingüístiques escollides, i establir comparacions entre les dues llengües dels nens bilingües que componen la mostra: el castellà com a llengua materna i el català com a llengua d'aprenentatge.

La descripció objectiva dels corpus analitzats i la quantificació i sistematització dels mateixos, ens poden aportar coneixements d'interès sobre el desenvolupament dels aprenentatges simultanis de les dues llengües, les interferències que es produeixen, els paral.lelismes que s'estableixen, la magnitud dels obstacles que es troben, la contaminació de les regles sintàctiques entre les llengües, en resum ajudar a tenir una comprensió més amplia de les lleis del llenguatge. Un millor coneixement de la evolució i desenvolupament lingüístics dels nostres alumnes fa possible una major

i millor formació i potencia noves estratègies en el sistema de organització del món mental infantil.

La meva pràctica professional, es veu freqüentment empobrida per la manca de coneixements que tenim de gran quantitat d'aspectes lingüístics. A les escoles, i possiblement a causa de la rigidesa dels programes oficials, es dona molta importància a aspectes com l'adquisició de l'ortografia o el coneixement de les parts de la oració, i molt poca a d'altres més globals: quin grau de maduració lingüística i competencial tenen els nens, quins elements gramaticals s'utilitzen amb prioritat, com es porta a terme la ampliació del vocabulari d'ús i de reconeixement, quins son els nivells acceptables de redundància i de repetició de frases, etc.

Després de fer el nostre anàlisis de prioritats, abordarem tota aquesta problemàtica des d'una perspectiva eminentment evolutiva i comparada de les variables escollides. Els resultats dels procés d'immersió, son els resultats de la comparació dels nivells de domini lingüístic entre les llengües catalana (d'aprenentatge) i castellana (materna).

El present treball consta de 6 Capitules i 1 Anexo. A traves d'els anirem tractant el següents aspectes:

En el capitulo 1, farem una introducció al tema que centri la problemàtica que volem abordar. Es delimiten les línies generals, la justificació i la utilitat teórico-lingüística i practico-pedagògica. En el Capitulo 2, es farà una revisió de la literatura científica prèvia i del estat actual de coneixements de la problemàtica que ens ocupa. En el Capitulo 3, parlarem del disseny i els objectius que intentarem cobrir. En el Capitulo 4, farem una descripció de la mostra, la metodologia del treball i com està organitzada la informació. En el capitulo 5, es dedicarà al anàlisis dels resultats obtinguts i tindrà 7 apartats. Cadascú serà tancat amb unes conclusions parcials, que finalment seran comentades de forma més global. El Capitulo 6, serà una síntesis de conclusions que donarà peu a una Discussió General amb la que posarem fi a aquest treball. En aquesta discussió estarà present la crítica dels aspectes més febles i vulnerables de la investigació.

El Annex, que se entrega en suport magnétic, un disc flexible de 3.25 polzades, els llistats dels vocables emesos en català i castellà respectivament, ordenats per ordre alfabètic.

1.1. PUNT DE PARTIDA.

La situació real de la nostra comunitat es que existeixen dues llengües oficials, que han de ser apresses al acabament de l'escolaritat obligatòria. En amples àrees del cinturó de Barcelona, la llengua materna majoritària es el castellà. Aquests alumnes han de aprendre el català i ho fan bàsicament a l'escola. Per això, s'ha posat en funcionament i en la actualitat està molt generalitzat l'aprenentatge en llengua catalana. Mitjançant una política d'immersió lingüística, els aprenentatges curriculars es fan en català. En síntesis, la immersió lingüística fa del català la llengua base de l'aprenentatge.

Aquest procés d'aprenentatge per immersió, ja ha començat fa uns anys, i no existeixen estudis que es fixen per objectius l'anàlisi del esmentat procés.

Seria bo plantejar investigacions capaces d'analitzar diversos aspectes inherents a la problemàtica que ens ocupa. El camí a recórrer no està exempt de obstacles, deficiències, interrogants, dubtes, etc. No es poden ignorar aquests fets si pretenen ser una mica científics i objectius. En la mesura que sigui possible hem d'intentar facilitar tot allò que es millor pels aprenentatges del nois i noies de la nostra comunitat bilingüe.

No està gens clar que l'aprenentatge de la segona llengua (L2 a partir d'ara) per immersió, sigui tan satisfactori com sembla. Al menys existeixen situacions on no es pot dir que tot vagi bé i que calen ser analitzades.

Encara no hem estudiat el comportament de allò que Lambert denomina bilingüisme additiu front al substractiu (Lambert, 1974), les implicacions que aquesta dicotomia implica i que poden arribar a un canvi de bilingüisme versus disglosia.

Tampoc saben res de si els nens que arriben a un nivell de competència lingüística molt baixa en L2 (sota un programa específic d'immersió) poden tenir pèrdues importants i significatives en la escolaritat posterior. De fet la interacció d'una segona llengua molt pobre, en situacions pedagògiques i entorns acadèmics cada vegada més simbòlics, pot estar per sota de la que desitgen com òptima per a potenciar i facilitar un progrés cognitiu adequat i uns resultats acadèmics positius. Cummings estudia el nivell de competència dels nens en les dues llengües i la relació d'aquesta variable amb l'aprenentatge bilingüe i el desenvolupament cognitiu (Cummings, 1979). Segons aquest autor, existeix un nivell mínim de competència lingüística que el nen bilingüe ha d'assolir per evitar que el bilingüisme provoqui l'aparició de dèficits cognitius importants. Tanmateix, hi ha un altre llindar més elevat que es necessari aconseguir per que cristalitzen les avantatges potencials del fet de ser bilingüe. Aquestes avantatges solament es manifesten quan es supera el llindar de competència en L1.

Altres autors (MacNamara, 1967), mostren la seva preocupació per la possible repercussió que pugui tenir el major temps de decodificació que utilitzen els subjectes quan reben la informació en la seva L2. Concretament MacNamara, es partidari de que la instrucció de determinats temes que requereixen grans quantitats de memòria a curt termini es portin a terme o bé en el idioma dominant dels nens o en el seu defecte amb un suport de material visual important. Amb aquesta mesura es pretén que el nivell de conceptualització pugui ressentir-se pel fet de la llengua d'aprenentatge no sigui la llengua dominant del alumne.

D'altre banda, el marc teòric dissenyat per Vigotsky i desenvolupat sobre tot per Luria, ens permet aportar una mica de llum a les no sempre clares relacions entre el llenguatge i el pensament (Vigotsky, 1973). Per la psicolingüística soviètica el llenguatge no es tan sols un reflex del pensament, o una forma de exteriorització-ho, sinó que al contrari es constitueix en un instrument per mig del qual el pensament pot assolir el rang de capacitat superior i per tant típicament humana. Des d'aquesta perspectiva el llenguatge es converteix en un segon sistema de senyals, arribant a ser un instrument necessari per el desenvolupament del pensament: "Pensem amb paraules". Això vol dir que, al menys en determinats supòsits, deficiències competencials lingüístiques, tant de tipus lèxic com semàntic o sintàctic, poden actuar com obstacles en els processos de pensament.

Tota aquesta línea de treball i investigació, es pot condensar en una frase d'un conegut lingüista: La lengua es una estructura informada de significació i pensar es manejar els signes de la llengua. (Benveniste, 1979).

Molt recentment, concretament la Conferència Internacional de Barcelona sobre Multilingüisme, sota la Presidència del Dr. Antoni Badia i Margarit, va aprovar un manifest[1] en que es demana als diferents Governs d'Europa que donin garantia de que tot ciutadà tingui dret a una escolarització en la seva llengua pròpia. Les conclusions del esmentat manifest fan èmfasi en la recomanació de que tots el nens tinguin la possibilitat de portar a terme el seu aprenentatge escolar en la seva llengua materna, sigui oficial o no.

El fil conductor argumental d'aquests autors de primera línea es la coincidencia en que el llenguatge no es un tema menor. Tampoc es un tema menor la llengua d'aprenentatge a l'escolaritat bàsica, com pensan per exemple les autoritats

---

[1] Manifest de Barcelona. Conferència Internacional sobre Plurilingüisme. Barcelona 6 de Abril de 1991.

académicas de la Generalitat de Catalunya[2]. De fet poden existir problemas i molt seriosos atribuibles a la generalització del Plan d'Immersió sense tenir en compte determinadas característiques socio-culturals dels nens i pares afectats.

Certament, existeix una gran dispersió teòrica en relació al tema de l'aprenentatge bilingüe. Sols així es pot explicar al menys part de lo que succedeix.

No n'hi ha un únic model d'aprenentatge bilingüe, existeixen molts, fins i tot contradictoris i oposats pel que fa a ritmes, prioritats, motivacions, supossits, utilització política, etc.

De tots, el escollit per la nostra administració es la immersió. La immersió consiteix en la exposició masiva en la escola del nen a la L2, amb la finalitat de afavorir molt aviat la seva presència.

El nostre model d'immersió al catalá té els seus antecedents en les experiències portades a terme en la regió francòfona de Quebec, al Canadà, on coexisteixen dues llengues oficialment: l'anglès i el francès[3]. La experiéncia de Quebec se'ns presenta com satisfactòria, i a la llum de les experiències fetes, ho es. Però son conscients de que aquest terreny es molt perillos. De les experiències sobre canvi de llengua llar-escola de Lambert, es dedueix que encara que els nens d'un nivell sòcio-cultural elevat tenen un suport a la pròpia llengua (L1) a la familia, els alumnes de nivells socio-culturals baixos no hi troben aquest suport. Com que normalment els nens que segeixen programes d'immersió provenen en general de nivells on només es desenvolupen determinades funcions del llenguatge i cap altres, cal que l'escola compensi el dèficit que això representa. Això vol dir que es fa necesari que en els programes d'immersió es prevegi el manteniment de la L1, ja que aquesta és una de les bases de l'èxit d'aquests programes.

Però les experiències de Quebec, no tenen ni el mateix disseny, ni la mateixa metodología, ni els mateixos objectius que tenim nosaltres. La nostra administració sembla que es partidaria d'una immersió, digeume, més dura: la L1 no té cap ajut dins del marc escolar fins uns quants anys després de

---

[2] Veure el contraste de les declaracions tan triunfalistes fetas per Joaquim Arenas, Cap d'Inspección d'Ensenyament, i d'altres molt més realistes de profesionals com Marta Mata o Maria Pla, en el suplement QUADERN, 380, pg.2, EL PAIS de 11 de Gener de 1990.

[3] A diferència dels Plans d'Immersió a la regió de Quebec, on els alumnes que han seguit la immersió al francés (L2) també estudian des de el primer moment al anglés (L1), a Catalunya, el castellá (L1) comença a introduïr-se 5 anys després de asistir a clase, es a dir en 3er. d'EGB.

anar a l'escola.

Per això nosaltres no tenim tan clara la generalització dels resultats positius que veus qualificades de la administració ens aportan sobre la immersió com mètode de ensenyament. Al menys no s'estàn complint els requisits que han fet de la immersió una forma d'aprenentatge exitosa o simplement satisfactòria.

Un problema es saber si el fet de que tots els aprenentatges es facin en català, mesura raonable per l'aprenentatge del català, no fa que alguns d'ells siguin més deficitaris que si es fan en la llengua materna dels nens. Perquè si fos així caldria plantejar-se el problema que es podria generar en un espai de temps mes o menys curt.

També es pot analitzar la evolució dels aprenentatges lingüístics i gramaticals del català a la llum de la immersió. Es de esperar que els resultats obtinguts siguin bons i acceptables. En cas contrari, caldria analitzar i discutir moltes premisses que semblen molt fixes.

Com es pot veure, la motivació per dur a terme aquesta investigació, és múltiple.

D'una banda la necessitat que tenim com a docents de mesurar els resultats obtinguts per l'aplicació d'un canvi pedagògic significatiu en l'ensenyament després d'un cert espai de temps. Ens referim als resultats sobre la millora de la competència lingüística del català dels alumnes que tenen aquesta com a segona llengua.

D'altre, ens proposem veure si l'augment tant de la qualitat com de la quantitat de lèxic català utilitzat, es significativament més elevat del que deun'en tenir els alumnes en el cas suposat de que portessin a terme el seu aprenentatge en la llengua materna.

Finalment seria bo, tant pel català com pels catalanoparlants potencials, poder comptar amb un Vocabulari Bàsic de Català que sigui un referent i un document en continua actualització en el marc de la nostra Comunitat. Ja fa anys que altres idiomes com l'anglès, el francès i recentment el castellà (que amplia i millora l'existent de Juilang-Chang), tenen els seus vocabularis bàsics respectius. En la mesura que poguéssim animar una tasca d'aquest tipus, o bé contribuir a la seva elaboració, ens sentiríem molt satisfets.

Treballs d'aquest tipus on es pugui consultar les paraules per a veure la freqüència que presenten en funció de determinades variables com (sexe, edat, llengua materna, zones, comarques, etc), ens ajuden a conèixer millor la llengua en general i l'ús particular que es fa d'ella.

En tot cas aquest seguiment lèxic que es proposa, adquireix una importància notable tenint en compte que el català coexisteix amb el castellà i les interferències son tantes i de tants tipus que caldria procedir a una enumeració de les mateixes. En aquest sentit i per si pot servir d'alguna utilitat, i sobre tot per aportar alguna millora, farem una

breu anàlisi descriptiva dels problemes ortogràfics (castellanismes i catalanismes) que hem trobat al llarg del nostre estudi.

No existeix cap raó per no començar un treball d'aquest tipus en l'idioma català. Al cap i a la fi no és una tasca ni tan complicada, ni tan costosa en comparació amb les aportacions directes que pot tenir un instrument com aquest en la pràctica escolar. Això per no assenyalar el component de prestigi, al menys dins del camp lingüístic, que adquireix un idioma que compta amb aquests instruments.

Tanmateix, la nostra motivació originària més profunda és la d'analitzar i contribuir a millorar els nivells de lèxic català usat en els diferents períodes de l'escolaritat. I això per diverses raons, entre les quals es poden citar:

.- La relativa poca importància que dins de l'ensenyament es dóna tradicionalment al lèxic, com aspecte fonamental del llenguatge. D'això en dóna fe el baix nombre de treballs que se centren en aquesta problemàtica.

.- La situació de contaminació del català per influència del castellà (també viceversa, però és un altre aspecte), derivada de la coexistència d'ambdues llengües. Comparativament la influència d'altres llengües estrangeres és molt menor, però no menyspreable. Seran aspectes a tenir en compte per tant, les errades de tipus ortogràfic.

.- La importància del lèxic des del punt de vista educatiu, donat que els dèficits de vocabulari, normalment es tradueixen en rendiments acadèmics molt baixos.

.- Afavorir l'adquisició del lèxic català en grups molt concrets d'alumnes (immigrants, estrangers, sords, deficients, etc). Els dèficits lingüístics de base d'aquests alumnes, fa necessari tenir en compte les paraules/vocables de major freqüència d'ús. Aquesta es una bona via per a la millor integració en la realitat en la qual estan inserits.

.- La construcció d'un instrument bàsic susceptible de ser utilitzat des de diferents àmbits i que incrementi els nivells d'eficàcia en l'ensenyament i l'ús del català.

Sóc sensible a la problemàtica que es deriva de la coexistència de dos idiomes en el camp escolar. El meu treball en la zona del cinturó industrial de Barcelona, eminentment castellà parlant, introdueix una sèrie d'interrogants objectius sobre les adquisicions lingüístiques de l'idioma català dels nostres alumnes a través dels anys. En relació a aquestes adquisicions, hi cap preguntar-se si realment els nostres alumnes estan aprenent el que nosaltres creiem que estan aprenent. Si el desfasament entre el lèxic que s'escolta, o el dels llibres, per exemple, i l'univers lèxic propi és excessiu, es corre el risc de provocar un dèficit conceptual que de perllongar-se en el temps seria irreversible.

No volem pels nostres alumnes un lèxic català reduït,

construït freqüentment a partir dels corresponents vocables castellans o disfressats o mutilats al final de la paraula. Pretenem que aquests alumnes que social i culturalment estan desfavorits, tinguin l'oportunitat d'incorporar la seva segona llengua d'una manera òptima, no basada en principis ni definicions apriorístiques, sinó en el seguiment del procés d'adquisició, en la crítica dels pressupòsits dels quals partim, de la metodologia que utilitzem i dels fins que aconseguim. No hi ha res tan negatiu com no qüestionar-se el propi treball. D'aquí a la fugida endavant només hi ha un pas. De vegades és molt important aturar-se, mirar enrera, veure la trajectòria i extreure les conseqüències que se'n deriven.

La metodología de recull de mostres escrites i de anàlisis de corpus d'aquesta investigació, podria ser un referent per a l'elaboració d'un instrument capas de reflexar el moviment que està tenint el lèxic català en un moment en que s'estàn incorporant parlants d'una manera tan masiva. Un Vocabulari Bàsic Català, extret de l'ús que es fa de l'idioma.

En funció del suport institucional, tan en recursos humans, com econòmics, seria procedent la recollida d'un mostrari lingüístic representatiu del català, que tingués en compte diferents tipus de variables (evolutives, econòmiques, regionals, culturals, dialectals, etc), que donaria com a resultat l'obtenció d'un corpus lingüístic (INPUT) que posteriorment s'analitzaria en funció de les variables establertes (OUTPUT) en la fase de disseny.

1.2. IMPORTANCIA DIDATICA DEL LEXIC.

El lèxic és un aspecte que podem considerar prioritari i fonamental del llenguatge, entès com a capacitat de comunicació més que com a disciplina curricular. Segons consta pels treballs duts a terme d'entre altres per la Psicolingüística Soviètica (Luria i Vigotsky), juga un paper determinant en els processos de pensament i desenvolupament de les estructures cognoscitives. De fet s'han comprovat nivells de correlació altament significatius (.81 segons Terman), entre la quantitat de lèxic usat i el desenvolupament mental.

Tradicionalment, la quantitat de lèxic, s'ha configurat com una variable potent en l'estimació de la intel·ligència, bé presa com un tot, bé com suma de factors. La competència lèxica, és una bona estimació de la capacitat intel·lectual.

Si això és així, es podria pensar que millorar el primer significa al menys, possibilitar un millor desenvolupament de la segona, però conèixer la quantitat i qualitat del lèxic dels nostres escolars, és previ a millorar-lo. La diagnosi és prèvia al pla d'actuació de millora lèxica entesa en termes no normatius, sinó reals. No n'hi ha prou en saber quina quantitat ha de ser el vocabulari mínim d'un nivell, sinó realment saber quin és.

Aquesta postura, lògicament implica una revisió dels criteris lingüístics en els quals es basa l'elaboració de llibres escolars. De vegades els nens no entenen res en els seus llibres, excepte els dibuixos a tot color. D'altres, tenen un nivell francament baix pels nens als qui van dirigits. Els dèficits de vocabulari tant des del punt de vista quantitatiu com qualitatiu, són en la pràctica la pedra angular de no pocs fracassos, manca de motivacions i apaties escolares.

Un cavall de batalla tradicional en referència al lèxic escrit, comú en tots els idiomes, ha estat l'ortografia. Recentment diferents llengües, de les quals n'és un exponent destacat la francesa, han obert una polèmica sobre la conveniència de mantenir la rigidesa o bé flexibilitzar la normativa ortogràfica vigent. En termes generals, donen suport al manteniment de les regles els lingüistes i els escriptors en front dels docents, més partidaris de facilitar la labor d'adquisició ortogràfica. Aquesta polèmica, ha arribat fins a nosaltres, si bé sembla que no ha tingut la mateixa amplitut que en altres llocs.

En tot cas, es pot fer un seguiment de tota aquesta problemàtica de l'ús i la norma lingüística seguint la trayectoria d'un grup de escriptors i lingüistes, que des de 1982, quan va apareixer el seu primer article a la revista Crònica, fins a la actualitat amb la publicació de **El barco fantasma (1982-1992)**, que recull una proposta d'ampliació del diccionari normatiu en més de 700 termes, ha tractar de posar en práctica el següent principi: un diccionari ha de descriure com és la llengua i no com hauria de ser. Estem parlant del

Grup d'Estudis Catalans, GEC[4].

[4] D'entre altres, autors com X. Pericay, M. Pla, O. Camps, J. Sellent, etc, son els impulsors del Grup d'Estudis Catalans. La seva finalitat es aconseguir que el català sigui una llengua més propera al carrer que a la autoritat lingüística.

1.3.-EL LEXIC COM ASPECTE RELEGAT.

Tradicionalment els docents sabem que en els darrers temps, potser dècades, el vocabulari no ha estat tractat acuradament en l'ensenyament. En part aquesta situació de marginalitat s'explica des de la gran complexitat i amplitud dels sistemes lèxics de qualsevol idioma, que generalment tendeixen a ser difícilment estructurables per la problemàtica de fixar uns límits de vocabulari que siguin estables per un determinat nivell o etapa educativa.

El ritme d'incorporació lèxica d'un idioma, que en aquests moments està creixent vertiginosament, implica un creixement continu de formes noves que és molt difícil mantenir actualitzat. A tot això s'ha d'afegir, per si no n'hi hagués prou, el creixement del nombre d'accepcions polisèmiques de determinats vocables. Cada dia augmenta el nombre de paraules que denominem la realitat i també la càrrega polisèmica de les mateixes.

Segons això, és bastant explicable que la majoria de les referències que es fan del vocabulari bàsic en els diferents nivells de l'Educació, siguin difuses i ambigües. En pocs llocs podem veure a què ens referim, de forma concreta, quan parlem per exemple del lèxic que ha de dominar un alumne de 1 d'EGB, o dit d'una altra manera, quin conjunt de vocables serien aconsellables per tal que un nen d'un determinat nivell pogués entendre i explicar la realitat que percep en funció del seu nivell de maduració conceptual.

La justificació de la necessitat de quantificar per nivells o d'altres variables els vocables utilitzats, neix de la impossibilitat de poder fixar-nos com a objectiu específic la totalitat del lèxic. D'aquesta impossibilitat s'en deriva la necessitat de jerarquitzar, prioritzar i seleccionar paraules o grups de paraules, però graduar la introducció de vocabulari a través de progressives fases evolutives i/o educatives, exigeix un coneixement previ de l'estat lèxic. A més, ha de ser un coneixement rigorós, exhaustiu i científic del mateix, basat en la realitat lingüística estudiada i no en la intuïció o en l'estimació subjectiva com malauradament passa més vegades de les desitjades.

I així arribem a allò que podem denominar l'objectiu bàsic d'aquesta investigació, que no és cap altre que el coneixement del lèxic real usat en els diferents moments evolutius que configuren la mostra.

Si cal i en posteriors fases, es pot desenvolupar programes de treball encaminats a introduir i ampliar, tant des del punt de vista qualitatiu com quantitatiu, els diferents nivells lèxics trobats, prenent com a punt de partida el que ja tenen els nens, i els increments lèxics anuals que experimenta la població escolar de forma natural, no induïda, o si es prefereix induïda assistemàticament dins del context social i escolar.

Contràriament al que passa en etapes anteriors a la

preescolaritat, que canalitzen la major part dels estudis centrats en Psicolingüística, el nostre interès es centra en el període escolar. Esperem contribuir a aportar algun coneixement lingüístic en aquesta etapa, que tan deficitària està d'aquest tipus de treballs.

2. LITERATURA PREVIA.

2.1. APORTACIONS GLOBALS I METODOLOGIQUES.

Citarem en aquest apartat els treballs de la literatura prèvia que han contribuït d'una forma global en el nostre. Més que extrapolacions i inferències directes, inviables per tractar-se d'idiomes diferents, hem de parlar d'aportacions de tipus metodològic, formal i d'enfocament analític. Amb l'ànim de no ser excessivament exhaustiu, citarem els següents.

En primer lloc, és obligat fer una referència a dos treballs que han suposat les primeres aproximacions a l'estudi lèxic del català. Es tracta de Borromba. Vocabulari bàsic il·lustrat i Vocabulari bàsic infantil, realitzats per Rosa Sensat. Ens han fet servei com a element de comparació i contrast dels resultats obtinguts per nosaltres, si bé existeixen diferències substancials en els mètodes utilitzats per a l'obtenció del lèxic.

El treball Funciones gramaticales en el habla infantil, dut a terme pel Consejo Superior de Enseñanza de Puerto Rico, 1960, que suposa per a l'idioma castellà una col·lecció de transcripcions de converses espontànies de parla infantil, anàlisis i observacions de molt interès. Es tracta d'una enorme mostra i conseqüentment un molt ampli corpus lingüístic de partida. El seu principal defecte, són les deficiències metodològiques en la recollida de material, que va ser agafat de viva veu, sense emmagatzemament en suport magnètic, per la qual cosa les transcripcions poden ser una font de variables no controlades al realitzar-les professionals amb criteris no sempre comuns.

Un altre treball també sobre lèxic de castellà, és el diccionari de Juilland-Chang Rodriguez, Frecuency Dictionary of Spanish Words, 1964. Més que per la seva qualitat i actualització, destaca per la falta d'altres treballs en aquesta línia. A diferència de l'anterior, que aporta una aproximació a l'ús oral de les paraules del castellà, parteix d'una massa lingüística escrita. En el nostre cas, l'aportació d'aquest treball, ha de ser valorada pel que fa exclusivament al tractament lèxic de la massa lingüística.

Però potser el més interessant dels treballs sobre vocabulari infantil en castellà, per la seva actualitat i la seva exhaustivitat, el Vocabulario básico en la EGB, dut terme pel MEC (1989). Aquest treball ens interessa sobre tot per les anàlisis de vocabulari, tant actiu com passiu, que es duen a terme a través dels 8 cursos de l'escolaritat bàsica, agrupats per nuclis temàtics i per cicles i pot constituir un pilar bàsic de posteriors estudis de lèxic infantil.

En relació a l'idioma francès, hem de citar Le français fondamental de l'Institut Pedagogique National. Es tracta d'una referència clàssica en els treballs de temàtica lexical.

Per altra banda, observem amb interès, una línia de treball que es proposa analitzar el llenguatge a través de paràmetres objectius. La quantificació de paraules d'una

producció lingüística, té valor per que a partir d'aquestes anàlisis es poden dur a terme certes interpretacions lèxiques. Aquest enfocament, inserit dins de la Teoria de la Informació, ha estat el primer intent d'estudiar el llenguatge com una magnitud física i objectiva, i com a tal poder ser descrita d'una manera formal i quantitativa. Autors com Zipf (1946), Herdan (1966), Muller (1968), etc, estudien diferents índex susceptibles de ser utilitzats com a elements que permeten la comparació de textos de forma assequible: variació intertextos que pot reflectir variacions de tipus estilístic, implicació lèxica del comportament de paraules amb una determinada freqüència per exemple 1, etc.

Aquest prisma d'anàlisi està molt connectat amb conceptes tals com "lingüística matemàtica", "estadística lexical", "lingüística computacional", "traducció automàtica", etc i un dels objectius prioritaris és la descripció quantitativa del lèxic, i partint d'això cotinuar treballant en el camp de la inferència. Recordem aquella frase lapidaria que sintetitza en bona mida aquest paradigma d'anàlisi: "La lingüística és la ciència estadística típica; els estadístics ho saben bé, la majoria dels lingüistes encara ho ignoren"[5].

---

[5] Pierre Giraud, *Problèmes et méthodes de la statistique linguistique*. 1960.

2.2. APORTACIONS LEXIQUES ESPECIFIQUES.

Existeixen diferents estudis centrats en observacions del comportament lèxic. Es tracta de treballs molt desconnectats, enfocats des de perspectives teòriques diferents i amb objectius puntuals i concrets. Això fa que no es pugui parlar amb propietat d'un corpus de coneixements lèxics consolidat, fiable i verificable. Més aviat tenim observacions i interpretacions de diferents dades i corpus lingüístics.

Amb tot, podem assenyalar dues línies bàsiques on agrupar els treballs de tipus lèxic. D'una part, l'estudi de les diferencies entre els codis oral i escrit. Per l'altra, l'estudi del comportament lèxic a través de l'edat. Freqüentment els mateixos treballs i autors estudien tot dos aspectes. Això fa que la tasca de recopilació d'aportacions de la literatura científica prèvia, no sigui una tasca fàcil.

Autors com Loban (1976), Hundscombe (1967), Vanoye (1973), Fawcet i Perkins (1980), etc. van estudiar les característiques diferencials intercodis. Per aquesta línia d'investigació la característica diferencial més important dels codis oral i escrit es que el primer conté més "mazes", o "garbles". Aquest concepte inclou tota la sèrie de paraules que en un moment determinat, perden el seu contingut semàntic.

També existeix acord en que la varietat de vocabulari, mesurat en número de verbs, adjectius, substantius i pronoms sempre és més gran en llenguatge escrit que en llenguatge oral. El nombre d'incorreccions gramaticals així com les redundàncies i repeticions en un text és més gran en llenguatge oral que en escrit. (O'Donnell, 1967 i Hunt, 1970).

Beudichon J., juntament amb Strook A. (1971), van estudiat el "décalage" evolutiu del lèxic, a partir de l'anàlisi d'alguns índex: número de verbs, adjectius, substantius, errors, repeticions, unitats independents, etc. Els seus resultats, manifesten tres diferències bàsiques entre els codis, i des d'una perspectiva evolutiva: Primer, en totes les edats, l'amplitud de vocabulari és superior en llenguatge escrit que en oral. Segon, els errors gramaticals i les repeticions produïdes, són substancialment superiors en llenguatge oral que en escrit. Tercer, la riquesa de les produccions dels codis (mitja de les unitats independents per producció), s'incrementa amb el desenvolupament evolutiu: la taxa obtinguda en llenguatge oral és notòriament més gran que en llenguatge escrit.

Diferents investigacions, manifesten que unes categories específiques de paraules, concretament els adjectius i els adverbis tenen un increment d'ús que és rellevant i significatiu en relació a l'edat, tan en termes absoluts com relatius (Gili Gaya, 1972). En aquest mateix sentit hem de citar les aportacions més rellevants obtingudes per Crystal (1980). Aquest autor va quantificar la proporció d'adverbis i de pronoms que existeixen en diferents tipus de construccions o de frases i va concloure que el 59 % de les frases tenen algun

adverbi, que aquest percentatge s'eleva al 66 %, si descomptem les frases en les que és impossible que existeixi, que l'adverbi apareix en el 28 % de les frases nominals i que considerats conjuntament adverbis i pronoms, aquests es presenten en el 57 % de les frases que es produeixen (sempre referit a l'idioma anglès).

L'ús de modificadors de nom tals com adjectius, participis, són progressivament més utilitzats (Hunt, 1965). També ho fan els adverbis que adquireixen progressivament més rellevància, cosa que indica una més gran maduresa en l'ús del llenguatge. (Quan parlem de categories gramaticals, ens basem en la taxonomia clàssica i tradicional, per ser més clara i menys dificultosa d'entendre).

En relació amb la quantificació de vocabulari, forçosament han de ser citats els treballs de Zipf (1949) sobre l'ús de les paraules. De fet la llei que porta el seu nom, Llei de Zipf, (encara que hi ha autors que l'atribueixen a Condon, i d'altres a Estuop), suposa l'inici d'una línia d'investigació basada en el plantejament estadístic de les freqüències dels vocables. En essència, la llei ens informa del següent: Considerat un text d'una determinada llengua, i després d'un recompte de les seves paraules, s'observa la diversitat d'ús de les mateixes. Unes apareixen més freqüentment que altres. Si classifiquem aquestes paraules en ordre decreixent a la seva freqüència d'aparició, (rang 1=paraula més freqüent, ...), es pot demostrar de forma experimental que el producte de la freqüència d'una paraula pel seu rang, és una constant. Aquesta línia de treball, ha tingut i sequeix tenint un considerable desenvolupament, potenciat per l'aparició de la lingüística computacional i l'anàlisi automàtica del llenguatge natural (Meya i Huber, 1986).

Dins d'aquesta corrent, són interessants els estudis sobre la longitud mitja de les paraules. Chebanov (1949), fou qui va començar aquesta línia d'investigació en llengües indoeuropees. Va Observar que la distribució de les paraules segons la seva longitud, sequeix la llei estadística de Poisson[5]. Segons aquest autor, la longitud de la paraula és capaç de diferenciar textos poètics, assajos o textos científics.

Més propers al català, són els estudis de Biermann a Barcelona sobre prosa literària en anglès (Hemingway, For Whom the Bell Tolls), en castellà (Baroja, Cesar o nada) i en català (Rodoreda, La Plaça del Diamant), han obtingut uns valors mitjans de 1.15, 1.89, i 1.53 síl·labes per paraula respectivament, citat per Marcus (1978).

S'ha estudiat la proporció que existeix entre les paraules que posseeixen un significat propi, independent del

[5] La distribució de Poisson es un cas límit de la distribució binomial, també es denominada llei dels successos rars.

context i les paraules que no el tenen. Seguint la terminologia usada per Stubbs (1986), denominem aquesta relació com Densitat Lexical, -per uns altres autors la densitat lèxica és la proporció de paraules/tipus per cada 100 paraules/ocurrència-, (Avila. 1972). La Densitat Lexical, l'obtenim per quocient entre el nombre de paraules lexicals i el total de paraules emeses. Es generalitzat considerar com a paraules amb un alt contingut semàntic els noms, adjectius, adverbis, verbs, i amb baix contingut semàntic o assemàntiques els articles, preposicions, conjuncions i interjeccions. Partint d'aquesta operativització, existeixen evidències de que el valor de la densitat lexical és un paràmetre diferencial interindividual i intertext, si bé com d'altres índexs han de ser analitzats amb cura.

La falta d'una estructura conceptual suficientment amplia, dificulta avenços significatius en aquest terreny. Els estudis realitzats recentment en llengua anglesa per autors com Ure (1971), Halliday (1985), Stubbs (1986), coincideixen en el fet que la llengua escrita generalment té una major densitat lexical que el discurs oral, si bé les desviacions són molt grans en funció de les mostres escollides per cadascun dels codis. Així per exemple, Stubbs (1986), ha trobat oscil·lacions entre la densitat lexical de dues mostres diferents de producció oral: comentaristes radiofònics -55%- i comunicacions telefòniques -44%-, superiors més de 10 punts. Com es veu les diferències són substancials. Tanmateix, i esperant que es facin més estudis, és preferible parlar de tendències, i esperar que posteriors treballs s'encarreguin de rectificar o ratificar les dades de Stubbs. Ure (1971), ha investigat aquest camp de treball, i ha arribat a la conclusió que els corpus de llenguatge escrit, tenen una densitat lèxica superior al 40%, amb un rang que oscil·la des del 37% al 57 %, mentre que la densitat lèxica del corpus oral és inferior al 40%, amb un rang d'oscil·lació entre 24% i el 43%. Per aquest autor, la línia divisòria entre els textos escrits i el discurs oral és aquell 40%. Això per l'idioma anglès, però com podem veure, aquestes dades no estan en completa sintonia amb els resultats obtinguts per M. Stubbs (1986), dues mostres de llenguatge oral ambdues superiors al 40%.

Altres treballs se centren en analitzar, dins de la frase, el comportament de les relacions establertes per les paraules modals. La modalitat es pot considerar com una entitat gramatical oposada al contingut proposicional. El seu estudi ens permet analitzar el posicionament del parlant respecte als fets del qui parla. Estudia les formes que adquireixen les manifestacions lingüístiques en funció de les distàncies que s'estableixen entre el productor del missatge i el missatge. El seu ús, generalment suposa un distanciament amb la producció i un relativisme o ambiqüitat amb la informació transmesa. Autors com Chafe (1986), estudiós del comportament dels verbs de modalitat: deure, preferir, pensar que,

haver de ..., han observat un major ús dels verbs modals progressivament amb l'edat i en produccions escrites.

En general i pel que fa referència als aspectes tant teòrics com metodològics no podem deixar de citar sobre tot el marc teòric proposat per Juilland i Chang-Rodriguez[7], citats abans.

A més a més de tots els aspectes concrets que hem comentat, hi ha molts treballs que tenen com a objectiu ho que per nosaltres pot ser una qüestió de fons: la situació bilingüe de la mostra. Dins d'aquest període escolar te lloc l'aprenentatge del català com a segona llengua.

Per autors com Fishman (1976), l'educació bilingüe és tot sistema d'ensenyament en el qual, en un moment variable i durant un temps i proporcions variables, consecutivament o simultàniament, es dona la instrucció almenys en dues llengües, de les quals una és la primera llengua de l'alumne.

Però la competència lingüística d'un bilingüe pot ser un element positiu (bilingüisme additiu) o negatiu (bilingüisme substractiu) segons, al menys en part, del tipus de competència que ha desenvolupat en L1 en el moment que comença l'exposició intensiva a la L2.

Autors com Cummings (1979), pensen que no sempre hi ha avantatges derivades de la situació bilingüe. Les avantatges només apareixen quan el nen té un nivell mínim de competència lingüística en L1[8]. Si no es aixi, la relació de la L2 amb l'escola es tornarà progressivament més simbòlica, i es situarà per sota de ho que sería desitjable com óptima per a la millora del seu desenvolupament lingüístic i acadèmic.

Per això ens sembla important conèixer, tant els ritmes d'adquisició com els obstacles que s'interposen per a assolir majors competències en L2 al llarg d'aquest període de sis anys.

En general, preguntes formulades fa molts anys encara no tenen resposta. No sabem quim mètode és el més racional i més segur per aprendre una llengua, ni quin és el moment més favorable per començar-ne l'aprenentatge, ni si l'estudi de totes dues llengües ha de ser simultani o successiu, etc[9]. Però en aquest treball no ens plantegem els citats interrogants, sinó d'altres més concrets, més operatius, més abordables i assequibles.

---

[7] Ens referims a *The Romans Languages and their Structures*, First Series SI, The Hague: Mouton, 1964, on es pot trobar de forma molt elaborada i exhaustiva els aspectes teòrics d'aquests tipus de treballs de recompte lèxic.

[8] Teoria del llindar.

[9] Conferència Internacional sobre Bilingüisme, Ginebra, 1928.

Al llarg d'aquest apartat, al marge de mencionar els problemes de fons de la situació bilingüe, hem tractat de reflectir el nivell de coneixements previ al nostre treball en relació a diferents aspectes del vocabulari: quantificació, caracterització, comportament i evolució. Sobretot, volem deixar constància de que el lèxic és un component lingüístic viu, flexible, amb entitat i comportament propi, que no es distribueix de forma homogènia i estàtica en els diferents corpus analitzats.

3. DISSENY.

3.1. ANALISI DEL PROBLEMA.

Hem revisat en el punt anterior els estudis realitzats sobre tot en anglès, francès i castellà, encaminats a obtenir una descripció sistemàtica del lèxic de diferents idiomes.

Tras un període de temps més o menys llarg de fer l'ensenyament en L2, convé estudiar el comportament de diferents variables gramaticals en L2 (llengua d'aprenentatge) i L1 (llengua materna).

Està molt clar que la política d'immersió lingüística des de una edat primerenca, objectivament afavoreix el coneixement d'aquesta llengua, però aquesta asseveració no impedeix plantejar d'altres interrogants.

- ¿Es deterioren aspectes concrets de la L1 (fluidesa, quantitat i qualitat de vocabulari, ortografia, etc.) per aquest fet?
- ¿Quines son les característiques de la L2 en el procés d'adquisició (obstacles, contaminacions e interferències amb L1)
- ¿Existeixen possibilitats de limitació estable i consolidada en els camps expressiu i comprensiu dels alumnes?
- ¿Els aprenentatges en els diferents curricula, es poden veure disminuïts per fer-los en una llengua no materna?
- ¿Quines son les diferències lèxiques en quantitat i qualitat que els nens utilitzen en la seva expressió escrita i en relació al català i castellà?.

Sembla que aquestes preguntes i d'altres que poden sortir, mereixen estudis ben elaborats amb criteris exclusivament científics, no ideològics, polítics o conjunturals.

A més a més d'intentar aportar alguna llum a aquests interrogants, ens proposem globalment fer fiable un mètode de treball, capaç de dur a terme la descripció quantitativa del lèxic català, per a mostres progressivament més extenses, que puqui contribuir a obtenir un Diccionari de Freqüències d'ús de les paraules. Com més extens i més bé elaborat estiqui, millor pot servir com a instrument al servei de l'idioma, en diferents àrees d'actuació. En tot cas l'aspecte que a nosaltres més ens interessa és lògicament el pedagògic.

3.2.-OBJECTIUS.

Després d'haver plantejat de forma global les pretensions d'aquest treball, passem ara a fixar i concretar els objectius que perseguim.

Podríem dir que l'objectiu general és analitzar el vocabulari bàsic infantil, reflex de l'univers conceptual dels alumnes.

El disseny del treball, farà possible comparar el vocabulari d'ús existent en català i en castellà, si volem la comparació objectiva entre la llengua de la classe i la llengua materna.

De manera més detallada, els objectius específics que pretenem omplir són per cadascú dels idiomes, d'entre altres:

- Obtenció del vocabulari bàsic.
- Riquesa del vocabulari bàsic.
- Evolució del lèxic.
- La longitud mitja de la paraula.
- Els solapaments lexicals.
- Nivells de redundància i densitat lèxica.
- Freqüència d'ús, polisèmia i longitud dels vocables.

A mida que anem desenvolupant aquests objectius, aniran sortint d'altres punts que pel seu interès, també seràn objecte d'estudi. Per citar tan sols alguns d'ells, la llei de Zipf en els textos, la importància de les paraules de Freqüència 1 (hapax en terminologia francesa), els errors de tipus ortogràfic, etc.

Resumint, pretenem aportar elements que permetin una reflexió realista sobre aquests problemes. Fer vàlid un sistema d'anàlisi lèxic, i contribuir en el seu cas a l'elaboració del lèxic usat partint de corpus lingüístics representatius i analitzar les diferències que existeixen entre les dues llengües dels nens, són les línies mestres que ens orienten.

4. METODOLOGIA.

4.1. LA MOSTRA.

La mostra de la qual hem partit en nostre estudi, està formada per 131 alumnes d'ambdós sexes d'un col legi públic situat en el cinturó industrial de Barcelona, concretament d'Hospitalet de Llobregat. De cada nivell acadèmic tenim al voltant de 20 subjectes, després de descomptar la mort experimental. El nivell socio-econòmic de les famílies dels nens és mig-baix. Un percentatge elevat, superior al 80%, de les famílies són emigrants, amb un temps de permanència a la ciutat superior a quinze anys, cosa que dóna una certa estabilitat i integració en el cos social del barri i de la localitat.

La majoria dels alumnes de la mostra, té el castellà com a llengua materna. L'edat dels subjectes cobreix una part de l'escolaritat bàsica, de 8 a 14 anys, és a dir dels cursos 3EGB a 8EGB, ambdós inclosos. Respecte a la variable sexe, la mostra està esbaixada a favor dels nens: 65% a 35%. Aquesta proporció reflecteix l'existent en el col legi per ambdós sexes.

Aquest col legi pertany al Pla d'Immersió Lingüística des de fa 5 o 6 anys, si bé ja abans d'aquest període ja es feia l'ensenyament en català. Això vol dir que la mostra analitzada ha rebut majoritàriament l'ensenyament en català.

De cadascun dels nens que van participar en aquesta investigació hem obtingut dues mostres escrites: una en català i d'altre en castellà. La llengua utilitzada en primer lloc s'ha balancejat.

Els registres van ser agafats, cap a final de curs. Concretament el mes de maig de l'any acadèmic 90-91.

Teòricament, tenim la pretensió de obtenir un corpus, millor dos corpus lingüístics, un en català i l'altre en castellà. Cadascú, tindrà al voltant de 30000 paraules[10].

---

[10] Els matemàtics es queixen de la manca de representativitat dels corpus linguistics. Les nostres cifres son una mica petites, però vàlides pel nostre treball amb un objectiu més exploratori que definitiu. No obstant, la manca de grandes masas de corpus es pot compensar amb mostreigs exhaustius.

4.2. PROCEDIMENT.

De cadascun dels alumnes de la mostra, hem obtingut dos fragments de la seva producció escrita[11]. La major part dels treballs d'aquest tipus, s'han elaborat amb corpus de llengua escrita per raons de facilitat de classificació dels textos, uniformitat, etc.

La fórmula exacta de la consigna que es donava, era: "Fes una redacció sobre el tema que vulguis, sobre el tema que en aquest moment se t'ocorri". Aquesta tasca seria duta a terme de forma col·lectiva. Prèviament, se'ls donaven instruccions per tal que no es preguntés en veu alta res, ja que havíem observat en una fase anterior del treball l'efecte de contagi que tenia el fet que un alumne preguntés si es podia escriure sobre un tema concret.

El temps que se'ls deixava, no era igual per a tots els alumnes tenint en compte que l'agilitat per escriure era diferent, tant a l'intergrup com a l'intragrup. Amb tot era suficient perquè la majoria podés desenvolupar folgadament el tema escollit.

En total tenim 131 registres de producció escrita. Hem volgut evitar que una major extensió de la redacció es tradueixi en una major quantitat de corpus. S'ha solicitat als nens que els textos tinguessin al voltant de 250 paraules, es a dir "un foli per una pàgina aproximadament".

Contràriament a d'altres autors[12], pensem que no es pot sobreestimar les produccions més llargues en detriment de les més curtes, ni identificar els registres obtinguts amb la competència lingüística dels alumnes ja que existeixen variables intermediàries com la planificació, l'agilitat i la motivació a la tasca entre d'altres, que podrien emmascarar resultats. El llenguatge de cada nen és un, comparable al dels altres, tant si és gran com si és petit, tant si és molt xerraire com si ho és poc. Per això, cadascú dels textos analitzats tenen un pes anàleg en el conjunt, una extensió aproximada, es a dir equilibrant l'extensió dels registres.

Abans d'introduir-los a l'ordinador pel seu processament, s'han codificat les paraules que presenten errors ortogràfics.

Amb l'ànim de diferenciar el que és un error ortogràfic intrínsec del vocable, i el que és una influència del castellà i per tant una limitació lèxica, hem fet una delimitació d'ambdós conceptes. Som conscients de la dificultat de la

---

[11] En paraules del estadístic-lèxic Vander Beke "Written discourse provides the only basis for marking a word list in a systematic and objetive manner", *French Word Book*, pg.11.

[12] Concretament ens referim a Müller, que es partidari de donar a les diferents parts del corpus densitats proporcionals al seu pes. Contrariament Juilland, amb qui estèm més d'acord, opta per mantenir un tamany similar en les parts del corpus.

tasca que proposen i assumim la responsabilitat dels errors que aquesta delimitació operativa pugui dur.

Així doncs considerem errada ortogràfica d'un vocable, qualsevol error en el mateix ja sigui per substitució, inversió o omissió de grafemes.

Considerem castellanismes aquells errors, que donen com a resultat un vocable igual o amb la mateixa base (arrel) que l'homòleg vocable castellà. Exceptuem dels castellanismes, aquells vocables amb error en un sol grafema, encara que el resultat d'aquest error sigui un vocable castellà.

Aquest ha estat l'apartat mes difícil de definir en termes de comprensió. L'intent de delimitació operativa del concepte castellanisme, es més complex que el fet de reconèixer-ho com a tal en textos catalans. Després d'un laboriós procés, no sempre exempt de contradiccions, hem arribat a definir operativament un castellanisme com el conjunt de grafemes que acompleixen les tres condicions següents:

1.- Aquest conjunt de grafemes, no existeix en català.

2.- Té mes d'un grafema equivocat.

3.- Té igual sema que la corresponent en castellà.

Si no és castellanisme per no assolir aquestes condicions, ho considerem una errada ortogràfica.

Amb tot aquest procés de codificació, volem evitar la dispersió d'ocurrències diferents degudes a errors ortogràfics, però que reflecteixen un mateix concepte, al temps que es té codificada aquesta informació per a posteriors treballs.

4.3. SEGMENTACIO D'UNITATS GRAFIQUES.

La pretensió del qualsevol diccionari d'aquest tipus es contar paraules semblants. Aquesta tasca aparentment senzilla, constitueix un dels problemes més aguts de la lexicometria actual. Decidir quan una unitat gràfica es pot considerar igual que un altre es més complicat que ho que aparentment sembla.

Tradicionalment aquesta unitat de segmentació a estat la paraula. Bé es cert que sota l'impuls de l'estructuralisme en l'estudi del morfema es va afavorir la idea de que la informació lèxica hauria de fer-se en termes de morfemes[13].

Front al morfema, la consideració de la paraula com unitat de segmentació ofereix les següents avantatges:

1.- Es més fàcil d'identificar (apareix separada per espais).

2.- Permet la explicació de les propietats distribucionals (formals) dels seus constituents (segons tingui flexió nominal o verbal).

La elaboració d'un diccionari de freqüències es pot realitzar a partir de una llista simple de paraules o bé d'una llista lematitzada.

La primera exigeix grans masses de paraules (items), i avui sembla descartada d'aquest tipus d'investigació[14]. La segona, les llistes lematitzades[15] són més assumides per la comunitat científica, si bé no deixen de tenir els seus inconvenients, sobretot pel que fa a

1.- La polisemia i homografia.

2.- Les seqüències de formes que funcionen com a un tot (he anat, si us plau, etc).

3.- Categorització de les paraules.

---

[13] Per a veure una exposició més amplia d'aquesta problemática, consultar la part introductòria de A. Juilland i P.M.H.Edwards, *The rumanian verb system*, The Hague: Mouton, 1971 i l'obra bibliogràfica de A.Juilland y A. Roceric, *The linguistic concept of word*,The Hague: Mouton, 1972.

[14] Es poden veure les reflexions de A. Lyne en "The problem of non-comparatibility of word-frequency" (*Proceedings Applied Contrastive Linguistic*, Vol. III, J.Quitgaarrd *et al.*, Heidelberg: J.Gross Verlag, 1972, pags. 468-481)

[15] Un diccionari lematitzat agrupa dins de l'unitat lèxica totes les formes flexives de la mateixa. Per ampliar aquest punt es pot consultar el *Recuento del vocabulario español* de Rodriguez Bou (pàgs. 649-650)

4.4. EL PROCESSAMENT DE LES DADES.

Per estudiar el comportament de la masses lingüístiques anteriorment delimitades, ens hem ajudat d'un programa informàtic, realitzat a tal efecte que s'ha encarregat d'extreure des dels textos, les paraules.

Posteriorment s'han realitzat les diferents fases de compte, comparacions, i d'altres càlculs de recompte i estadístics.

Com a resultat, hem obtingut per de cada llengua, una base de dades de vocables[16], en format DBASE III+.

El número de vocables diferents que hem obtingut per cada llengua està per sobre de 2000, i en ells es reflexen tots els conceptes diferents que apareixen en les 131 redaccions, després de diferents tasques de filtrat i despulla. Les formes d'aquests vocables són la reducció més simplificada que hem pogut obtenir:

- Els verbs es presenten en infinitiu. Les formes composades han estat desdoblades (auxiliar + verb). La major part dels participis ha estat considerat com adjectiu.
- Els noms, adjectius, pronoms, etc, s'han reduït a singular i masculí.
- Les formes polisèmiques no han estat desdoblades.
- Desdoblament de formes gràfiques on apareixen guió o apòstrofe amb articles i pronoms.
- Eliminació dels noms propis i abreviatures.
- Eliminació dels cardinals, però no dels ordinals.

De cada element o paraula, indiquem 8 camps o punts d'informació que són, a part del número d'ordre corresponent, el concepte, la freqüència amb què apareix als textos en cada curs estudiat i la freqüència amb què apareix al total dels textos. Els noms dels camps son: Paraula, FR3, FR4, FR5, FR6, FR7, FR8 i FRT respectivament.

---

[16] El vocable es el representant lèxic, si es vol la unitat semàntica de totes les paraules o cadenes gràfiques amb la mateixa lema. La literatura saxona, utilitza el conceptes type i tokem, anàlecs als nostres vocable i paraule respectivament.

5.- RESULTATS.

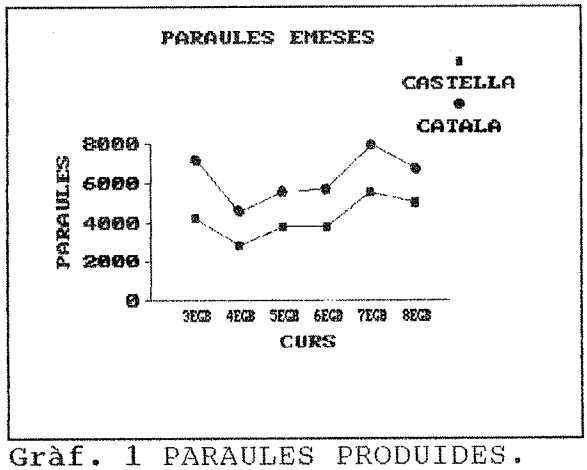
5.1.- VOCABULARI: PARAULES I VOCABLES.

5.1.1.- COMPARACIO DELS CORPUS.

El primer anàlisis que hem fet ha estat la quantificació de les masses de corpus per cadascú dels idiomes.

Recordem que la consigna que donaven als nois era que escrigueren sobre un tema qualsevol al voltant d'un foli. Això tant en castellà com en català. La elecció de la primera llengua estava balancejada, i en la segona redacció es mantenia constant el tema de la primera.

Doncs bé. La quantitat de corpus que els nens utilitzaven en català es significativament superior que en castellà. El Gràf. 1 presenta els corpus de cada llengua des de una perspectiva evolutiva.



Gràf. 1 PARAULES PRODUIDES.

Com podem observar, en català hem obtingut globalment un corpus significativament mes gran que en castellà (aproximadament un 30%). Aquest percentatge es manifesta com una constant al llarg de tot el període estudiat.

A la vista d'aquests resultats, podem fer alguns comentaris.

En primer lloc, comparant els corpus produïts per totes dues llengües, es veu clar que existeix una major capacitat, domini o fluïdesa del català escrit. Es a dir, en general els textos escrits en català son més llargs que els escrits en castellà. Recordem que els alumnes tenien total llibertat per a escriure la quantitat de text que li semblava.

La explicació més fàcil que podríem fer seria atribuir aquest resultat a la manca de domini del castellà escrit. Però aquesta explicació no té massa consistència. Es cert que l'aprenentatge del procés lecto-escriptor es fa en llengua catalana i això implica un major coneixement de la llengua escrita en català en els primers nivells, però això no explica el mateix fenomen uns anys després. No té cap sentit pensar això, si no suposem que aquesta inferioritat es manté al llarg de tota l'escolaritat, lo qual es en tot cas una hipòtesis de

treball no un asseveració demostrada.

Un altre possible explicació podria ser la manca general de motivació, costum o hàbit per escriure castellà en aquests alumnes encara que tinguessin les capacitats suficients. Possiblement cap de totes dues explicacions sigui certa al cent per cent, i cadascuna aporti una part de veritat.

5.1.2.- ELS VOCABLES OBTINGUTS.

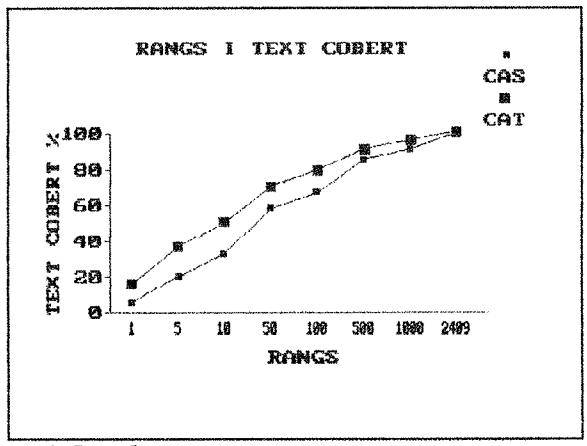
Si bé el corpus lingüístic en català es superior al corpus castellà, una primera aproximació al anàlisis dels vocables obtinguts en els respectius corpus, aparentment entra en contradicció a les observacions del punt anterior.

En efecte, després de portar a terme totes les transformacions necessàries mitjançant un procés de lematització, en obtingut pel català un numero inferior de vocables (2115) que en castellà (2409).

Els vocables, com ja dèiem en altres punts, son els representants dels corpus emesos. Podíem dir que son una mida de la capacitat de denominació de conceptes que existeixen dins dels nens. Una major quantitat i/o qualitat de lèxic, de vocables, es tradueix indefectiblement en una major facilitat per expressar les idees i sentiments, en general, el mon del pensament.

Si agafem les llistes de vocables de les dues llengües i les ordenem per freqüència d'ús de manera que els més usats apareguin abans que els que tenen poc ús, obtindrem una llista ordenada per rangs. Tindrem lògicament tants rangs com vocables en la llista.

Si comparem la quantitat de text que cobreixen determinats rangs en la llista, podem veure que el percentatge de text cobert pels mateix rangs es superior en català que en castellà. En el Gràf. 2 tenim una representació del que estem diem.



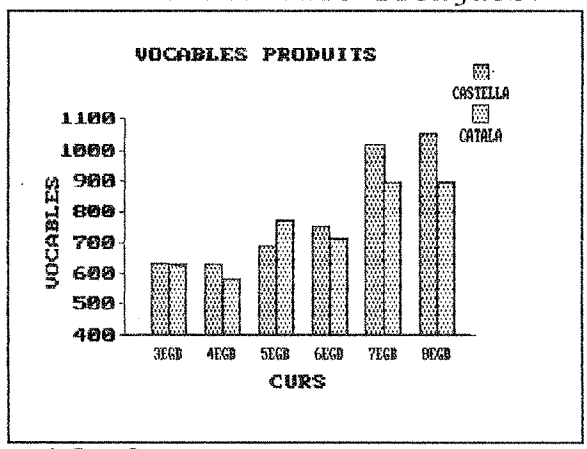
Gràf. 2. RANGS I TEXT COBERT.

En efecte, la primera paraula de la llista de vocables de català cobreix el 15% del text produït. En castellà la primera

paraula cobreix el 6%. Les 5 primeres paraules de cada llista expliquen el 37 % i el 21% para català i castellà respectivament, les 10 primeres cobreixen el 51% i el 33%, etc.

Aquests resultats manifesten que el conjunt de vocables catalans de molta freqüència, son molt més utilitzats que els més freqüents en castellà. Fins i tot es podria dir que aquest grup de vocables, moltes vegades adquireixen un comportament proper al que tenen quan s'utilitzen en llenguatge oral, es a dir com a "comodín". En aquest sentit, els vocables catalans d'aquest grup tindrien un major comportament com a "comodins" que el corresponents en castellà.

D'altre banda, el número de vocables que es produeixen en cadascú dels sis curs de la mostra, sempre es major en castellà que en català (amb excepció de 5è d'EGB). En el Gràf. 3 podem veure el número de vocables que s'utilitzen en cada curs i en cadascuna de totes dues llengües.



Gràf. 3. VOCABLES PRODUITS.

Resta dir que com es lògic, i molt conegut per tothom, aquest gràfic reflexa d'una manera global l'ús progressiu de més vocables a mida que la edat/escolaritat es major. Es a dir quan més grans son els nens més gran es el número de vocables que utilitzen.

En síntesis allò que podem veure en el present apartat, son dos fets que a primera vista semblen contradictoris.
En primer lloc, la mostra produeix una quantitat de corpus significativament mes gran en català que en castellà, tant aïlladament en cada com al conjunt de la mostra.
Contràriament, quan analitzem el comportament dels vocables (formes simplificades de cadascuna de les unitats del corpus), es posa de manifest que es més petit en català que en castellà, tant globalment com curs per curs amb l'única excepció de 5é.

Això ens permet fer els següents comentaris.

Primer, que en termes generals el corpus català es més redundant que el produït en castellà.

Segon, si els contrastem, veurem que la sobre-utilització

dels vocables de molta freqüència en català es superior a la que es produeix en castellà.

Tercer, com que hi ha menys vocables, o es lèxic es més pobre, la comunicació no té el nivell òptim de precisió que caldria.

Quart, i com a conseqüència de l'anterior, la càrrega polisèmica dels vocables es superior a la desitjada.

5.1.3.- UNA ESTIMACIÓ DE LA REDUNDÀNCIA LÈXICA.

Podem estudiar la Redundància Lèxica en termes de proporció de vocables produïts en relació al total de paraules (cadenes gràfiques separades per un espai) emeses.
¿Quants vocables obtenim per cada 100 paraules produïdes?
¿Quantes paraules de cada 100 eliminem en el procés de depuració per l'obtenció de vocables?
¿Quantes paraules de corpus necessitem per obtenir un vocable depurat, un concepte?

Les paraules eliminades en aquesta tasca de transformació constitueixen la redundància lèxica. Una manera senzilla d'operativitzar la redundància, si bé no la única, podria ser dividir el total de paraules emeses pel total de vocables obtinguts, es a dir:

Taula I Redundància lèxica.

$$RL = V/P$$

on: RL = Redundancia lèxica.
V = Total de Vocables obtinguts.
P = Total de Paraules emeses.

Amb aquesta fórmula obtindríem el total de paraules necessàries en qualsevol fragment del corpus, per obtenir un vocable.

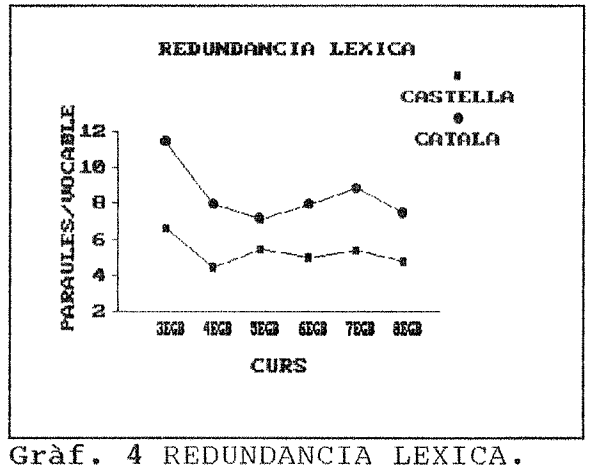
La Redundància Lèxica pot configurar-se com una variable evolutiva i diferencial interllengües.

Podem veure la comparació de les redundàncies per curs i en cada llengua en el Gràf. 4.

En efecte, aquest gràfic ens presenta dos fets molts clars. Evolutivament, les taxes de Redundància lèxica tras una baixada important (sobre tot en català) de 3é a 4rt d'EGB, podem dir que es mantenen gairebé estables al llarg del període estudiat. Seria un altre tema d'estudi aquesta estabilitat de la redundància. Pel que fa als nivells de redundància comparada interllengües, es pot veure que aquesta es menor en castellà que en català d'una forma molt sensible.

Però, sembla que no n'hi ha explicacions massa coherents a aquests dos fenòmens. ¿Per què les diferències que s'observen entre la redundància en català i en castellà, no es van

fent progressivament més petites? ¿Fins a quin punt es pot pensar que es mantindren amb el temps uns nivells de redundància per la llengua catalana tan superiors als que es mantenen en castellà?. Aquests i d'altres interrogants plantejats no tenen resposta dins d'aquest treball. Deixem per un altre investigació l'anàlisi del comportament evolutiu i diferencial de la estabilitat de la redundància lèxica mesurat de la manera indicada.



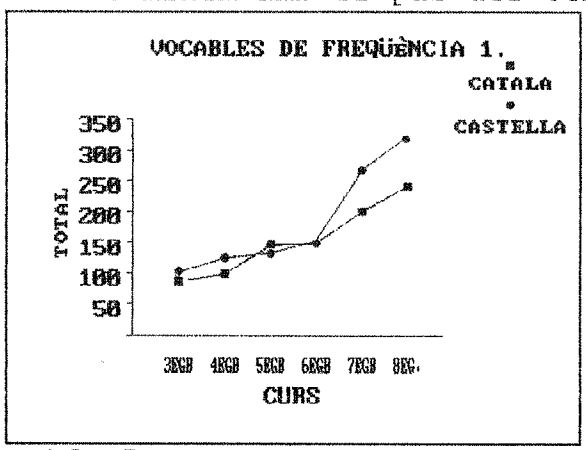
Gràf. 4 REDUNDANCIA LEXICA.

5.2. VOCABLES MÉS I MENYS ACTIUS.

5.2.1. VOCABLES DE FREQÜENCIA 1.

El conjunt de vocables que apareix una sola vegada en un determinat vocabulari -F1-, ha estat un paràmetre molt utilitzat dins del camp lèxic (Girauld, 1960). De fet ho primer que fa un vocable abans d'incorporar-se a un lèxic habitual, es tenir freqüència 1, restant en aquesta posició més temps quant més petit sigui el seu ús.

Segons això es podria dir que els vocables F1 ens donen la imatge congelada d'una situació dinàmica de flux d'assimilació lèxica, de trànsit de vocabulari passiu a actiu. El coneixement d'aquest paràmetre dins de un determinat grup lingüístic, ens indica la velocitat d'incorporació de lèxic nou en aquest grup. Quan més gran sigui major es augment lèxic i viceversa. Lògicament la magnitud F1 es inversament proporcional a la quantitat de lèxic que es posseeix, per la qual cosa tendeix a fer-se mínim amb el pas del temps.



Gràf. 5 VOC. DE FREQUENCIA 1

Aquest gràfic mostra el comportament dels grups de vocables de freqüència 1 al llarg del període, tant en català com en castellà. Podem veure que:

Primer. En totes dues llengües es pot apreciar un increment evolutiu del grup de paraules de freqüència 1. Es a dir en el període estudiat té lloc un creixement global del lèxic, cosa que ja sabíem.

Segon. El català presenta un creixement que es més bé sostingut i discret, sense grans saltes.

Tercer. En castellà es podem observar dues fases. Fins a 6è tenim una línia de creixement suau. A partir de aquest moment, el creixement lèxic es torna més accelerat.

Quart. Comparativament, en castellà s'observen tasses més elevades de vocables de freqüència 1 que en català, si bé en el últim tram aquesta diferència es fa més patent.

Resumint, en castellà els ritmes d'incorporació lèxica de vocables nous es produeixen de forma més ràpida que en català, sobre tot en els cursos 7è i 8è.

Aquest aspecte està relacionat amb el major número de vocables que s'obtenen en castellà encara que el corpus de partida sigui més gran en català. També guarda perfecta coherència amb les explicacions i interpretacions donades en apartats anteriors d'aquest treball.

5.2.2. ELS 50 VOCABLES MES USATS.

Quan relacionàvem els rangs amb la quantitat de text produït, mostràvem que les 50 primeres paraules cobrien més del 50% dels corpus -Gràf. **2**-. Aquestes llistes representen més d'una de cada dues paraules produïdes. Per això es important veure si existeixen diferències qualitatives entre de les dues llistes que ens pugin donar informació sobre les característiques estructurals d'aquests grups de vocables de molt ús.

En la Taula que s'adjunta s'ofereix el llistat de les 50 vocables més usats en català i en castellà.

Taula II Vocables més freqüents.

NUM	CATALA	CASTELLA	NUM	CATALA	CASTELLA
1	el	el	26	nen	muy
2	a	y	27	se	para
3	ser	un	28	dia	del
4	anar	que	29	més	comer
5	de	de	30	aquest	niño
6	i	a	31	veure	ver
7	en	se	32	ell	cuando
8	un	ser	33	com	me
9	que	en	34	perquè	más
10	haver	ir	35	si	otro
11	li	haber	36	ja	si
12	al	le	37	meu	casa
13	no	no	38	voler	llamar
14	hi	al	39	passar	dar
15	fer	lo	40	casa	porque
16	dir	por	41	poder	mi
17	molt	su	42	arribar	mucho
18	tot	decir	43	altre	coger
19	per	con	44	tenir	querer
20	estar	todo	45	noi	salir
21	del	pero	46	pare	vez
22	seu	estar	47	això	llegar
23	amb	hacer	48	jo	ya
24	però	tener	49	després	poder
25	quan	día	50	sol	o

Com veiem les 50 paraules més usades en cada llengua pràcticament son les mateixes. Moltes d'elles inclús coincideixen en el rang.

No se han trobat en general diferencies significatives en la composició de categories gramaticals d'aquesta mostra de paraules més freqüentes.

De fet, aquest resultat es previsible ja que partim dels mateixos nois que fan dos mostres de la seva producció lingüística. Es a dir partim dels mateixos universos conceptuals en les dues produccions. Es fàcil obtenir resultats semblants, al menys en el grup de vocables de més ús.

Tan sols hem d'assenyalar un parell de diferències: una de tipus qualitatiu i l'altre quantitativa.

La primera consisteix en el número de verbs que apareixen en aquest grup de paraules. En català tenim 12 verbs i en castella 16, ho que representa un 25 % menyes de verbs en català que en castellà. No sabem amb exactitud la profunditat d'aquesta observació, però si posteriors estudis fossin capaços de demostrar la estabilitat i el manteniment d'aquesta diferencia, estaríem davant d'un fenomen de major acció i dinamització del discurs castellà al menys escrit.

Es possible que la clau d'aquesta interpretació estigues basada en les diferències entre la conjugació de verbs en català i en castellà, o bé que els nens, per inseguretat, utilitzaren mecanismes d'evitació de la conjugació amb la finalitat de no cometre errades.

Penso que aquests aspectes mereixen ser abordats de forma més local i específica a com ho fem aquí. Nosaltres, dins d'aquest treball general, tan sols volem deixar-lo assenyalat per si pot servir per treballs futurs.

La segona diferència fa referència a la mitjà aritmètica del dos conjunts. El grup de vocables més usats en català té un promig de 532 molt superior al de vocables més usats en castellà que es de 293. Això confirma ho que es deia en el punt 5.2. i 5.3. sobre la quantitat de corpus coberts pel mateixos rangs i sobre la redundància dels textos. Pel que fa a les dispersions d'aquestes distribucions no se han trobat diferències significatives.

5.3.- LONGITUD MITJA DELS VOCABLES.

L'estudi de la variable longitud mitja dels vocables, a estat iniciat per Chevanov en llengües indoeuropees. Pel que es refereix a les llengües romances, existeix un estudi de Biermann, que compara la longitud mitja dels idiomes castellà, català i anglès.

Com síntesis dels estudis portats a terme por Chevanov en 127 llengües indoeuropees diferents, es pot extreure una conclusió important. En la llengua literària, i més concretament la poètica, s'utilitzen generalment paraules més curtes que en els textos científics.[17]

En la llengua poètica tenen gran ús les paraules del fon lèxic bàsic de la llengua. Aquest fon lèxic bàsic està constituït per paraules de gran circulació, i de longitud molt petita.

Els textos científics normalment utilitzen vocables i conceptes especials, que no pertanyen al fons lèxic. Com que la creació de vocables del fons lèxic està sotmesa a les combinacions fonemàticas coherents amb les lleis de la llengua, els nous vocables científics han de ser forçosament més llargs.

En el llenguatge habitual, es necessari que les informacions siguin transmeses ràpidament. Per això el fons bàsic està format espontàniament per vocables curts[18].

Existeix una forta relació entre economia i lingüística, que fa que els vocables més usats siguin més curts que els menys usats. L'ús s'encarrega de disminuir la longitud de la paraula.

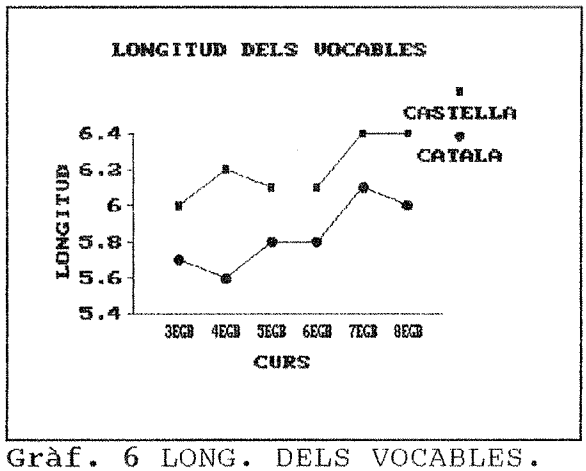
Normalment la longitud dels vocables s'expressa en número de sil.labes. Nosaltres, i donat el comportament més o menys estable de la sil.laba, expressarem la longitud dels vocables en termes de número de lletres per la raó de ser més senzill i més precís.

El següent gràfic mostra l'evolució de la longitud mitja dels vocables usats en cada una de les dues llengües i des de la perspectiva evolutiva d'aquest fragment de la escolaritat obligatòria.

---

[17] Per profunditzar en les diferències entre lenguatge literari i científic, es poden consultar els treballs de Fucks: Teoría de Fucks "Mathematical theory of word-formation", *Information Theory, Third Londom Symposium* (Colim, ed), Londres, 1956.

[18] Un cas que posa de manifest allò de que estén parlant, es a dir la tendencia a la economia lingüística, es la recent incorporació per part de la Real Academia de la Lengua com a formes correctas en castellà, vocables com a zoo i tele que susbtitueixen respectivament a zoològico i televisión.



Gràf. 6 LONG. DELS VOCABLES.

Aquest gràfic ens permet veure la longitud dels vocables utilitzats al llarg del període i per cadascú de totes dues llengües.

Es pot apreciar que els vocables empleats en català son en general més curts que els utilitzats en castellà. Això es coherent amb els estudis de Biermann, citats abans. Recordem que els seus resultats donaven una longitud mitja dels vocables catalans de 1.53 sil.labes i dels castellans de 1.89.

També es pot veure que existeix una lleugera tendència a incrementar la longitud amb el pas del temps. Hi ha una major longitud mitja del vocables a mida que augmenta a edat i/o escolaritat.

Sabem que des de la perspectiva de la Teoria de la Informació[19], la longitud dels vocables no es una variable neutra. Podem dir que dins d'una llengua, una major longitud dels vocables es correspon amb una major quantitat de informació transmesa.

En efecte. Si agafem un text que conté una paraula amagada, pot ser que el context ens aporti la informació que no tenim o bé que es generi una certa quantitat d'incertitud. Ho que passi dependrà sobretot de la freqüència d'ús de la paraula en qüestió. Si la paraula te molta freqüència d'us tenim moltes més probabilitats d'endevinar-la que si és una paraula poc freqüent. Però com que les paraules de molt ús son normalment curtes i les de freqüència baixa més llargues, existeix una relació directa entre la longitud de les paraules i la quantitat d'informació que aporten (o la incertitud que la seva presència pot eliminar).

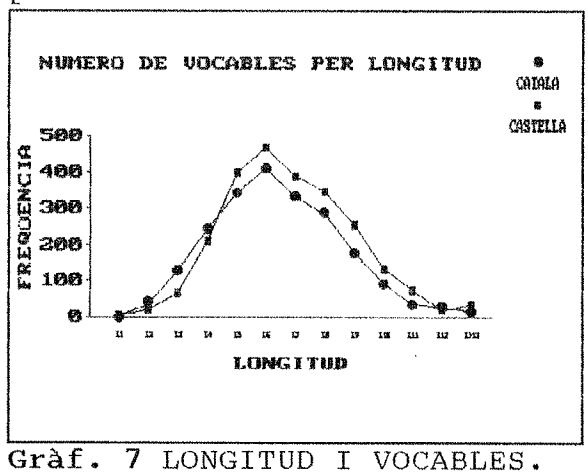
Segons això, amb ho que nosaltres estem d'acord, els nens

---

[19] Shannon en "A mathematical theory of comunication", *Bell System Tecnical Journal*, 27, 1948, exposa els principis de la Teoria de la Informació. No obstant, hi ha un precursor: Hartley (1928) que ja parlava de la "quantitat de informació".

més grans aporten més quantitat de informació en els seus escrits que el nens més petits. I com que els vocables que s'utilitzen expressen el mon de les idees, els pensaments, els coneixements, podem dir que el mon dels conceptes es més elaborat, més complex i ric a mida que els nens son més grans.

La informació d'aquest paràmetre (longitud dels vocables) pel que fa a la comparació interllengües, està fora del nostre camp d'anàlisis, però es podem trobar interessants observacions, com les que es mostren en el següent cuadre, i que podem ser ampliades en diferents treballs dintre de la lingüística comparada.



Gràf. 7 LONGITUD I VOCABLES.

Veiem que, en coherència de allò que venim parlant al llarg de tot aquest apartat, els vocables obtinguts en la mostra de castellà presenten freqüències més petites que en la mostra de català pels vocables curts o molt curts. Contràriament, pels vocables llargs, la mostra de català presenta freqüències inferiors que les obtingudes en castellà.
Fent una mica d'extrapolació, podem dir que quan els nens s'expressen en català utilitcen vocables més cur

5.4.- ELS SOLAPAMENTS LEXICALS.

Podem classificar el conjunt dels vocables que s'utilitzen en cadascú dels sis moments estudiats, en grups segons la proporció que tinguin de vocables comuns a tots el cursos o específics de cada curs.

En aquest apartat de solapaments lexicals estudiarem les característiques que posseeixen els vocables en funció del seu ús més o menys comú al llarg del període estudiat. Podem formar tres grups de vocables:

1.- Els vocables comuns utilitzats els sis anys.

2.- Els vocables semiespecífics, utilitzats per més d'un curs però no per tots.

3.- Els vocables d'ús específic en qualsevol curs.

El anàlisis del comportament d'aquests tres grups al llarg del període estudiat, ens pot informar sobre aspectes importants com les possibles zones de solapament lexical, la efectivitat dels increments dels vocabularis actius, l'estimació del lèxic que s'aprèn (se compren -lèxic passiu-, però no se posa en funcionament en la expressió -lèxic actiu-, etc.

En primer lloc, el grup de vocables comuns a tots els cursos està format per 169 vocables en català i 160 en castellà.

Aquest grup te unes característiques diferencials en relació els vocables específics:

a.- Es tracta de paraules de longitud curta.

b.- Tenen una freqüència d'ús molt elevada.

c.- Majoritàriament tenen funció de relació, asemàntica.

En segon lloc, els vocables semicomuns, compartits per més d'un curs (però no per tots). En termes generals, podem assenyalar que el número de vocables comuns que tenen dos cursos es proporcional a la seva proximitat cronològica o acadèmica. El grup de vocables semicomuns se comporta de forma anàloga en castellà i català tant des d'el punt de vista quantitatiu, com de la dispersió mitja intercursos.

En tercer lloc tenim els vocables específics de cada curs. A diferència dels grups anteriors, aquest grup es caracteritza per:

a.- Generalment no presenten freqüències elevades, sinó més bé discretament baixes.

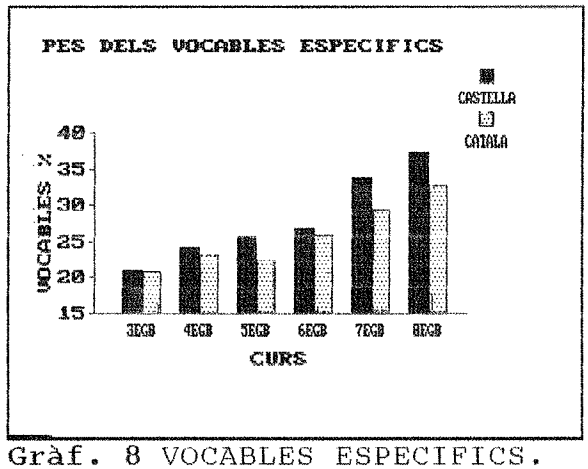
b.- Gramaticalment aquest grup està constituït per noms, adjectius i verbs.

ca.- Evolutivament, els vocables específics perden materialitat i concreció i guanyen en immaterialitat i abstracció.

d.- L'escolaritat es veu reflectida en aquest grup de vocables. Un número significatiu de vocables específics, son importats des d'els continguts de les distintes

materials curriculars[20].

En la gràfica que presentem a continuació mostrem el comportament evolutiu dels vocables específics en les llengües estudiades.



Gràf. 8 VOCABLES ESPECIFICS.

El gràfic ens permet fer dues observacions interessants:

1.- Els vocables específics tenen major pes dins del total de vocables produïts, en castellà que en català. Això en tots els cursos sense cap excepció.

2.- Des d'el punt de vista evolutiu podem veure que el grup de vocables específics també té més pes quan els cursos con més elevats.

Es a dir en la llengua materna, el grup de vocables específics té un creixement evolutiu més fort que en la llengua d'aprenentatge, que també creix però d'una forma més discreta. Sembla doncs que una bona part dels nous vocables introduïts pels curriculums, són traduïts al castellà abans de passar a formar part del fons lèxic. Segons això existeix una transformació o una traducció automàtica inherent a la tasca d'emmagatzement.

Aquesta interpretació no està gaire al.lunyada de la polèmica existent en torn als models de emmagazement i tipus d'organització de la informació. Sense cap interés de entrar dins d'aquesta problemàtica tan complexa, direm que a nosoltres ens sembla que els nostres resultats s'orienteriem més cap a la concepció de que el processament de la informació es

---

[20] Consultar **El Lenguaje Oral y Escrito en la EGB.** Juan Antonio Cordero Alonso. Tesis Doctoral. 1990. Universitat de Barcelona, inèdita.

fa básicament mitjançant un magatzem únic[21]. Aportem la date per si pot servir per ajudar a reforçar aquest o una altre d'aquests aspectes.

---

[21] Podem profunditzar en el tema del tipus d'emmagazatment en Mariano Yela, **Estudios sobre inteligencia y lenguaje.** Pirámide. Madrid. 1987. Pgs. 218-243.

5.5.- POLISEMIA, US I LONGITUD DELS VOCABLES.

En el punt 5.2. hem parlat de la relació directa existent entre la longitud que tenien els vocables i la quantitat de informació que aportàven.

En aquest punt ens disposem a estudiar la relació que existeix entre la càrrega polisèmica dels vocables, la freqüència d'ús i la longitud dels mateixos.

La manera que proposem per definir operativament la càrrega polisèmica dels vocables es assignar-les un número que representi el total de significats diferents (o el número de contextos diferents en los que es poden utilitzar) que troben en qualsevol diccionari estandaritzat de la llengua. Lògicament aquesta tasca no pot portar-se a terme de forma individual sinó globalment, per a grups d'intervals.

Concretament nosaltres hem utilitzat el Diccionari Barcanova de la Llengua[22] per català i el Diccionario General Ilustrado de la Lengua Española[23] per castellà.

Partint del conjunt de vocables emesos ordenats de manera decreixent segons la seves freqüències, formarem 3 grups de 15 paraules cadascú (aleatorització dels 15 primers vocables del millar). En el primer grup contindrà els 15 primers vocables, en el segon grup estaran el vocables situats entre els rangs 1001 i 1015 ambdós inclosos i en el tercer, els vocables del 2001 a 2015.

Els resultats es presenten en la següent Taula.

Taula III Longitud, polisemia i freqüència d'ús. Mitjas.

RANGS	LONGITUD		POLISEMIA		FREQUENCIA	
	CASTE	CATAL	CASTE	CATAL	CASTE	CATAL
1/15	2.2	2.4	7.7	9.3	654	1401
1001/15	7.6	6.7	5.4	3.6	2	2
2001/15	7.7	6.7	3.7	2.4	1	1

[22] Dirigit per Xavier Romeu i Juvé i amb pròleg de Albert Jané. Primera edicció: 1985. Barcanova. Barcelona.

[23] Dirigit per Manuel Alvar Ezquerra i prologat per Samuel Gili Gaya. Primera Edición: Sept. 1987. Biblograf. Barcelona

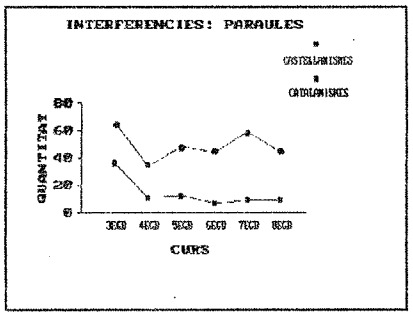
5.6.- INTERFERENCIES LEXIQUES.

Com dèiem en el punt 4.2., abans d'introduir els textos en l'ordinador es feia un treball de codificació de les paraules amb algun tipus d'error. Es a dir, es codificaven les faltes d'ortografia.

No pretenen en aquest apartat analitzar les errades que han aparegut en els nostres textos, sinó mostrar l'evolució de dos tipus concrets d'error: els castellanismes (interferències del castellà en el català) i els catalanismes (substitucions de paraules castellanes per les corresponents catalanes).

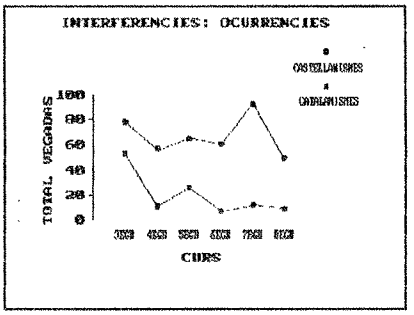
Es poden consultar la operatització del conceptes castellanisme i catalanisme en el apartat 4 del present treball.

Els resultats que hem trobat es reflexen en el Gràf. 9 i 10. El primer mostra el total de paraules diferents (formes) que es produeixen com interferències lèxiques. El segon, mostra el número total de vegades que han estat produïts catalanismes o castellanismes en els textos.



Gràf. 9 PARAULES.

De fet els gràfics mostren la mateixa realitat si bé amb diferents unitats de mida. Per tant, farem comentaris comuns que seran aplicables a tot dos. Podem observar que:



Gràf. 10 OCURRENCIES.

1.- Els catalanismes estàn per sobre dels castellanismes, i sembla que es mantenen sense cap decreixement estable ni consolidat.

2.- Els castellanismes mostren una tendència a la baixa al llarg del període, essent molt forta de 3è a 4art.

Sembla que aquests resultats no son del tot satisfactoris. Els nens de la mostra han portat i estan portat a terme la escolaritat obligatoria en català. La pregunta que ens hem de fer es ¿perquè hi ha una persistència tant forta de castellanismes en llenguatge escrit, quan la llengua escolar (on es més gran la component escrita que la oral) es el català?

Penso que ja no es pot atribuir tota la presència dels castellanismes a la pressió del castella fora de l'aula com es feia fa 10 o 15 anys. Tampoc es una explicació convincent del fenomen dir que la llengua familiar té tanta força que es capaç de produir tantes interferències dins l'altre llengua.

Es més lògic pensar que si es fan castellanismes es pura i simplement per que existeixen llacunes lèxiques o gramaticals importants en el català. Es per això, i amb la finalitat de que el flux comunicacional no resti entrecortat, que

s'agafen vocables de l'altre llengua, on si existeixen els recursos lèxics.

Pel que fa referència als catalanismes, quantitativament molt per sota dels castellanismes, la majoria dels etiquetats com tal són molt utilitzats per la comunitat castellanoparlant, al menys per l'entorn social i familiar de la mostra.

En resum, la presència d'interferències d'una llengua en l'altre, posa de manifest les deficiències puntuals o en el seu cas estructurals del domini d'una o de les dues llengües.

Si existeixen llacunes lèxiques i semàntiques importants en català, que sembla que si, cal preguntar-se on es troben les causes de que tal situació, cal plantejar-se si la metodologia didàctica per l'aprenentatge del català es la millor de les possibles, cal analitzar d'altres métods més eficients i menys contraproduents d'aprenentatge de la L2, cal analitzar fins a que punt aquestes deficiències estàn relacionades amb la manca de domini de la L1 com pensa per cert Cummings, cal criticar que el temp dedicat a la L1 no sigui significativament més grand que el dedicat a qualsevol llengua extrangera, cal plantejarse la llengua d'aprenentatge, cal que els aprenentatges de l'escola no s'oblidin de la llengua materna del nen, i que en base a aquesta comenci l'aprenentatge de la L2. Cal, en fí, balancejar si els sacrificis que imposa la immersió primerenca i sostinguda són compensats pels resultats obtinguts.

Després d'un temps de immersió, s'han d'estudiar els resultats de la forma més objectiva possible. Sembla que la situació no es tan optimista con es pensava i que per si mateixa no es capaç de solucionar els problemes que apareixen quan els nens castellanoparlants utilitzen la llengua catalana. La situació no dona més de si. Aquesta línea d'actuació sembla esgotada.

5.7.- CONCLUSIONS.

En el apartat de Resultats, hem exposat les dades que hem trobat al llarg de la investigació. Recordem que es tracta d'una mostra de nens i nenes d'un Col.legi Públic de L'Hospitalet, que majoritàriament tenen com a llengua materna el castellà i que han fet la major part dels seus aprenentatges escolars en català.

Presentem a continuació les reflexions generades per les dades obtingudes al llarg del present treball.

En primer lloc hem vist que els nens de la mostra han produït una quantitat més gran de text en català que en castellà. Durant els primers anys de l'estudi (3é i 4art) aquest fet pot explicar-se a partir de la manca de domini escrit que encara tenen els nens, però com que aquesta tendència no es corregeix posteriorment, sembla que sigui més degut a la manca de motivació per escriure en castellà que a la manca de domini escrit.

Però a pesar d'això, quan comparen el numero de vocables o formes depurades en les dues llengües, trobem que en castellà s'obté una major quantitat de formes que en català. Es a dir, encara que en català existeix un major corpus lingüístic de partida, els vocables o formes diferents obtingudes són menys. Recordem que estem parlant de tests escrits. D'això es dedueix que en català existeixen uns nivells de redundància superior als que es presenten en castellà, o dit d'un altre manera, en el català escrit de la mostra es acceptablement profús però poc ric en matisos. Donat que l'univers conceptual dels alumnes es més ampli del que es reflexa en el lèxic català obtingut, com ens demostra l'existència de més vocables en castellà, la utilització del català escrit pot convertir-se en un obstacle per explicar els pensaments, les idees, els sentiments, etc. Al menys pot ser un obstacle per aconseguir un determinat nivell de precisió semàntica. I pensem que a llarg termini, això pot ser perillós.

En segon lloc, un anàlisi del comportament de F1, o conjunt de vocables emesos una sola vegada, posa de manifest que tant al llarg de cadascú del sis moments estudiats com del període global, el paràmetre F1 es més fort en castellà que en català. Aquest paràmetre, recordem, mostrava el ritme d'increment de vocabulari en un moment determinat, o sigui el conjunt de vocables que estan començant a utilitzar-se. Que vol dir això? Com pot ser que el ritme de creixement del lèxic català sigui més petit que el del castellà, encara que l'escolarització (font fonamental de nous conceptes i termes) sigui portada a terme en català?. Sembla que el procés de adquisició de nous conceptes en la L2 no es tan lineal com es pensa. L'aprenentatge es fa en català, però a la vista dels resultats podem dir que la interiorització del concepte es fa en castellà de tal forma que quan el nen parla o escriu en la seva L1, aquests termes tenen més probabilitat de sortir que quan

s'expressen en la seva L2. O dit d'una altra manera: el rendiment expressiu dels nens, al menys escrit, es més elevat quan s'expressen en la seva llengua materna, encara que sigui de temes escolars o acadèmics, desenvolupats en català. Anàlogament, el pes dels vocables específics en la L1 es més gran que en L2. Però com que els vocables específics (molt a prop dels continguts acadèmics) s'introdueixen en català i es manifesten prioritariament en castellà, sembla que existeix una traducció immediata a la L1, de d'on l'utilitzen.

En tercer lloc, les nostres observacions confirmen les fetes per Biermann en el sentit que les mitjas de longitud de paraula son més petits en català que en castellà. Però dit això, caldria verificar si aquesta observació pot amagar d'altres comportaments que existeixen encara que de forma latent. Ens referim a la relació entre la quantitat de lletres i la quantitat de informació transmesa.

Tal com posa de manifest la Teoria de la informació, pot ser que una part d'aquesta longitud més petita estigui relacionada amb una menor quantitat d'informació que transmeten els textos en català. Com hem explicat en el seu moment, els vocables més llargs, en general aporten més quantitat d'informació que els més curts, es a dir eliminen més incertitud. De tota manera, no aprofundirem en aquest punt, tant sols ho assenyalem. Sembla que això que diem es possible. Els nostres anàlisis apunten en aquesta direcció, i no descartem abordar aquesta problemàtica en posteriors treballs.

En quart lloc, destaquem que no existeixen diferències profundes en els vocables més utilitzats en cada llengua. Es tracta d'una analogia sistemàtica pel que fa a aspectes semàntics (els vocables més freqüents s'utilitzen amb el mateix significat i amb els mateixos matisos), grafèmics (molts vocables son físicament els mateixos en totes dues llengües), gramaticals (tenim les mateixes categories gramaticals i quasi bé els mateix vocables per cadascuna d'elles), etc. En certa manera això, al menys part d'això, es podia suposar a priori: son els mateixos nens els que fan els dos corpus, existeix una analogia profunda entre el català i el castellà com llengües derivades del llatí, etc. No obstant això, seria bo estudiar les analogies que es presenten en aquest grup de vocables de molta freqüència, per si poden ser un indicador de com pot evolucionar el català a la llum de la incorporació masiva de castellanoparlants que te lloc en la actualitat.

Per últim, comentar el tema de les interferències interllengües que existeixen.

Es normal que en una situació de coexistència de dues llengües, hi hagi una certa quantitat de interferències entre elles. Es més fàcil que aquestes es presentin a nivell oral, però també apareixen a nivell escrit si bé menys perquè la escriptura està sotmesa a un major grau de formalització. De tota manera, el nostre treball mostra una descompensació de

les interferències produïdes, una falta de simetria. Hem trobat que les interferències del castellà en el català son majors que les que es produeixen del català al castellà, la qüal cosa seria normal si parléssim de llenguatge oral. Que es doni aquesta situació en textos escrits, quan els nens han fet la major part de la escolaritat en català, i tenint en compte que l'aprenentatge acadèmic té una fort component de llenguatge escrit, es lo que sembla menys normal. Quina explicació es pot donar a aquest fenomen? En aquest moment no estem en condicions de poder donar una resposta a aquesta pregunta, però no es massa arriscat pensar que existeixen, diguem "asistencies tècniques" d'una llengua a l'altre quan surten deficiències o mancances en la llengua activa. En aquest sentit, es possible que n'hi hagi hagut més interès institucional per que el català es generalitzi a tothom i tot arreu, que per donar uns nivells de qualitat més acceptable que els que es mostren. En tot cas aquesta ha estat una decisió política que no critiquem. Si es pot dir que tal decisió porta de forma inherent una aproximació lenta però sistemàtica de les dues llengües, cosa molt sana d'altre banda.

7.- DISCUSSIÓ GENERAL.

Hem vist al llarg d'aquest treball tot un conjunt de fets i dades sobre el comportament lingüístic de dos corpus el català i el castellà, produïts per la mateixa mostra de nens i nenes: un Col.legi Públic de L'Hospitalet de Llobregat. Aquesta mostra es caracteritza per tenir majoritàriament com llengua materna el castellà i com a llengua d'aprenentatge el català.

Per començar hem de fer una autocrítica de la nostra investigació.

La primera es refereix si la quantitat de subjectes mostrejats es significatiu i permet fer generalitzacions o bé derivacions a d'altres mostres més grans. Lògicament si la quantitat de subjectes no es considera suficientment ample, els corpus obtinguts a partir d'ella també seran insuficients. De sempre s'ha plantejat una objecció genèrica a totes les investigacions que s'originen a partir de mostreigs: la representativitat de la mostra. Nosaltres no estem d'acord en que tot mostreig sigui deficient per se. Pensem, això si, que la qualitat del mostreig està relacionada amb la qualitat dels resultats obtinguts. Des d'aquesta perspectiva hem treballat, conscients de les avantatges i inconvenients, de les aportacions i deficiències que ens aporta la metodologia estadística. Dins d'aquesta línia podem parlar d'un determinat poder de generalització dels resultats. La mostra de la que hem partir ha estat de 131 nens i nenes, i les conclusions que es deriven de l'estudi creiem que poden ser generalitzables a d'altres zones arreu de Barcelona i del seu cinturó.

Un altre crítica que ens fem es n'hi han temes que han sorgit al llarg del treball i no han pogut ser analitzats en profunditat. En el punt 3, parlaven la problemàtica i dels interrogants que ens plantejàvem. No tots han estat estudiats com correspon a la seva importància, però també es cert que el nostre disseny de recollida de dades permet un cert enfoc i no d'altres. Parlar, per exemple, dels perjudicis d'aprenentatge que puguin tenir nens amb la seva llengua materna poc assolida i malament estructurada, suposa d'altres enfocs, altres dissenys, altres corpus. Tant d'aquest com d'altres punts deixem la porta oberta per fer altres treballs més específics.

Però bé, passem a comentar d'una manera global, de síntesi, les reflexions que ens permet els resultats obtinguts.

Per començar direm que si l'únic objectiu que persegueix la línia educativa de la nostra Administració es aprofitar l'escolaritat bàsica per fer que els nens castellanoparlants comprenguin i siguin capaços de expressar-se en català, hem de dir que el objectiu està amplament cobert. Es la dir la política d'immersió portada a terme els últims anys ha estat eficient.

Ara bé si creiem que ha de ser l'Administració la que ha de estar al servei de la ciutadania i no al contrari, l'objec-

tiu de que abans parlàvem es manifestament insuficient. Una cosa es que tots els que viuen en una comunitat bilingüe tinguin el deure de conèixer els dos idiomes de la comunitat i utilitzar el que vulguin en cada moment i una altre diferent es que els nens castellanoparlants hagin de fer el seu aprenentatge en català.

Suposo que l'administració no ha tingut temps, ni diners ni ganes de analitzar, i no pressuposem res, quines son els efectes negatius o si volem mancances derivats de la immersió. I si s'han fet aquests estudis, i si son realistes, científics, objectius, verificables, on estan? On es tracten aquestes problemàtiques, o es que no existeixen? Coneixem als científics que treballen per que la realitat sigui lo més semblant a allò que ells mateixos pensen que hauria de ser: de seus treballs s'aprofiten les parts que ajuden a explicar allò que es vol, però tot es aprofitable? Que mes coses es troben i sembla que s'amaguin? Es lícit desconfiar d'aquest tipus de ciència? Encara més, això es ciència?

El que proposem es començar a analitzar el procés d'una forma global, sense cap prejudici, valorant encerts i defectes, facilitant canvis si podem optimitzar els resultats educatius en termes de guany objectiu pels nens i no sols si son capaços de parlar una mica en català al final de la escolaritat.

S'hauria d'analitzar si en algun moment concret, sobre tot els primers cursos d'escolaritat obligatòria, n'hi han hagut problemes de comprensió de conceptes pel fet de ser explicats en una llengua sobre la que el nen no té massa competència, o bé si en determinats conceptes, tradicionalment difícils, s'ha incrementat la seva dificultat intrínseca amb altre de tipus lingüístic que neix de la manca de comprensió i precisió semàntica, o bé si existeixen nens que quan comencen l'escolaritat als 3 anys, resten marcats per que no entenen el seu professor/a, o com viuen els més petits el canvi de començar anar a l'escola si a això s'afegeix la manca de comunicació al menys en una direcció (professor/a alumne/a), o si estem mutilant i no afavorint la capacitat expressiva dels nens certament una mica tímids, vergonyosos, etc, que no parlen per por de fer el ridícul.

Tot això i moltes coses més han de ser tingudes en compte abans de parlar de l'èxit del procés d'immersió. Després de balancejar allò que hem aconseguit i allò que hem deixat d'aconseguir inclús de les deficiències generades, podem parlar d'una manera neutral dels resultats. No abans.

Sembla que el nostre model en matèria de immersió lingüística i que tan satisfactòriament s'ha portat a terme en Canadà, més concretament a Quebec, part de pressuposts diferents als nostres: els pares amb nivells de cultura molt elevats, desitjaven que el seus fils fessin en la escola l'aprenentatge de la L2. Eren els pares els que demanaven els canvis en la llengua d'aprenentatge. Això vol dir que dins de

d'aquestes famílies la L2 era viscuda com una cosa pròpia que volien tenir per a els. La nostra situació no es la mateixa. En el nostre cas tothom sap que existeixen dues llengües oficials, tothom es partidari de que es coneixen bé totes dues, però no tothom veu bé que la llengua d'aprenentatge sigui un altre que la seva llengua materna. Tampoc s'ha donat la oportunitat de que la gent volguesi això. S'ha fet per decret, en un moment de poques demandes i pressió socials, ningú ha dit res, però es tracta d'un tema molt delicat i sembla que encara no està tancat.

D'altre banda, l'experiència de Quebec, no es tan rígida ni tan radical com la nostra. A Quebec, la immersió primerenca dels nens en la L2, no oblida mai la existència de la L1. No tan sols als nens se'ls hi dona classe, se'ls hi parla en la seva llengua materna, també es fa un seguiment exhaustiu de com evoluciona la L1 amb el pas del temps. Les autoritats de Quebec basant-se en autors tan significatius com Cummings, pensen encertadament que l'èxit de la competència en la L2 està en relació directa a com es desenvolupe la L1. No es pot assolir uns nivells de competència important en L2 si L1 està sota mínims. Nosaltres no se si pensem d'aquesta manera o no, però està clar que no fent lo mateix que ells. El temps s'encarregarà de dir-nos que s'ha fet malament, però per llavors per molta gent serà massa tard.

I això es tot. Si les dades que presentem poden introduir o facilitar una reflexió sobre allò que estem fent a les escoles referent al tema de les llengües des de fa anys, sense prejudicis, sense posicionaments previs, el treball haurà estat positiu.

8. BIBLIOGRAFIA

Avila, R y otros (1972): **Léxico infantil en Mexico.** Mejico. Trillas

Badia i Margarit, A. (1962): **Gramàtica catalana.** Madrid. Gredos.

Beaudichon-Strook. (1971): "Etude du 'décalage' ent langage oral et langage ècrit d'après l'analyse de quelques indexs significatifs". **Enfance**, 71, 4-5, 353-375.

Benveniste, E. (1979): **Problemas de Linguística General I.** Madrid: Siglo XXI

Benveniste, E. (1979): **Problemas de Linguística General II.** Madrid: Siglo XXI.

Bustos, E. (1967): "Un nuevo recuento de vocabulario español". **Filología Moderna.** Madrid, 25-26 (1966-1967).

Crystal D. (1980): "Negleeted grammatical factors in convers National English". En Greenbaum et al. (Eds). **Studies in English Linguistics for Randolph Quirk.** London: Longman.

Chapman i Czerniewska (1978): **Reading from Process to Practice.** London.

Cheshire, J. (1982): "Dialect features and linguistic conflic in schools". **Educational Review**, 34: 53-67.

Cummings, J. (1979): "Linguistic interdependence and educational development of bilingual children". **Review of Educational Research**, 49 (2), 222-251. Existeix traducció al castellà en **Infancia y Aprendizaje**, 1983, 21 (1), 37-62.

Ducrot, Oswald. (1986): **El decir y lo dicho.** Buenos Aires: Paidos.

Escarpit, R. (1973): **L'écrit et la communication.** Paris: Presses Universitaries de France.

Fabra, P. (1966): **Diccionari general de llengua catalana.** Barcelona.

Fawcett i Perkins. (1980): **Child language transcripts 6-12**, vols I-IV. Pontypridd: Polytecnich of Wales.

Fontanillo M.E. (1986): **Diccionario de Linguística.** Madrid: Anaya.

Garcia Hoz, V. (1977): **Estudios experimentales sobre Vocabulario.** Madrid: C.S.I.C.

Gili Gaya, S. (1972): **Estudios de Lenguaje Infantil.** Barcelona: Vox.

Giraud, P. (1960): **Problémes et méthodes de la statistique linguistique.** París: Presses Universitaries de France.

Herdan,G.(1964): **Quantitative Linguistics** London:Butterworths.

Hunsdcombe, R. (1967): " The written language of twelve years old children. Nuffield foreing languages teaching materials project", reports and accasional papers, n. 26-28. London: **The Huffields Foundation.**

Jordana, C.A. (1968): **El català i el castellà comparats.** Barcelona. Nova edició.

Juiland-Chang. (1964): **Frecuency Dictionary of Spanish Words.** La Haya: Mouton.

Justicia J.,F. (1985): **El Vocabulario Usual del Niño en CI y CM de la EGB.** Granada: Instituto Ciencies Educacion.
Karmiloff-Smith, A. (1979): **A functional appoach to chidren language.** Cambridge: Cambridge University Press.
Katz i Fodor (1963): "The structure of a semantic theory". En Fodor i Katz (Eds). **The structure of language: Readings in the philosophy of language.** Englewood Cliffs, NJ, Prentice-Hall.
Knapp, K. (1982): **La comunicación no verbal.** Barcelona: Paidos.
Lambert, W.E. (1974): "Culture and language as factors in learning and education", en F. Aboud i R.D. Meade (eds.) **Cultural factors in learning,** Bellingham, Western Washington State College.
Lara, L.F. (1979): **Investigaciones linguísticas en lexicografía.** Mexico.
Lentin, L. (1971): "Recherche sur l'acquisition des structures linguistiques chez l'enfant entre 3 et 7 ans", **Etudes de Linguistique appliquée,** 4, pgs. 7-50.
Loban, W. (1963): "Language of elementary scholl children". Research Report, n 1. Illinois: **National council of teachers of english.**
Loban, W. (1976): **Language development: Kindergarten though grade twelve.**Urban III. National council of teachers of english.
Lorenzo, D. (1980): **El Vocabulario y la Ortografía de nuestros alumnos.** Madrid: Cincel-Kapelusz.
Lyons, John. (1985): **Introducción a la Linguística Teórica.** Barcelona: Teide.
MacNamara, J. (1966): **Bilinguism and primery education,** Edimburgo, Edinburg University Press.
MacNamara, J. (1967): "The bilinguals' linguistic performance: A psychological over view", **Jour of Social Issues,** 23, pgs. 58-77.
Mandelbrot, B. (1965): **Information theory and psycho-linguistics.** New York: Wolman-Nagel.
Marcus, S... (1978): **Introducción a la lingüística matemàtica.** Barcelona: Teide.
M.E.C. (1989): **Vocabulario bàsico en la EGB.** Madrid: Espasa Calpe.
Meya, M. i Huber W. (1986): **Lingüística computacional.** Barcelona: Teide.
Muller, Ch. (1973): **Estadística lingüística.** Madrid: Gredos.
Rodríguez Bou, I. (1952): **Recuento de Vocabulario Español.** Puerto Rico.
Rosa Sensat. (1976): **Borromba. Vocabulari bàsic illustrar.** Barcelona: La Galera.
Rosa Sensat. (1978): **Vocabulario bàsico infantil.** Barcelona: Bibliograf.
Smith-Deirce. (1979): **La linguística moderna.** Barcelona: Anagrama.

Stubbs, M. (1986): "Lexical density: a tecnique and some findings". En M. Coulthard (ed.): **Talking about Texts.** Birmingham: English Language Research.

Stubbs, M, (1986): **Language development, lexical competence an nuclear vocabulary.** E, K. Durkin ed Language Development in the Shool Years. London: Croom Helm.

Stubbs, M. (1987): **Anàlisis del discurso.** Madrid: Alianza Editorial.

Ure, J. (1971): "Lexical density an register differentiation". En G. Perren i T. (Eds). **Applications of Linguistics.** Cambridge: Cambridge University Press.

Vanoye, j. (1973): **Expression Communication.** Paris: Armand Collin.

Vigotsky, L. (1973): **Pensamiento y lenguaje.** Buenos Aires: Pléyade.

Yela, M (1987): **Estudios sobre inteligencia y lenguaje.** Madrid. Pirámide.

Zipf, G.K. (1949): **Human behavior and the principle of least effort.** Cambridge: Mass