

Le redoublement total dans les langues d'Europe

Thomas Stolz

Université de Brême, Allemagne

Table des matières

- Résumé
- Littérature
- Définitions
- L'universalité douteuse
- L'Europe
- Deux études quantitatives
- La fin

Resumé

Ven. 15/01/10
14h00
Bât. D
Salle 311

Pr. Thomas Stoltz (U. de Brême)
Le redoublement total dans
les langues d'Europe

Présentation

Les langues européennes sont-elles dépourvues d'un procédé structural quasi universel comme le redoublement total ? La réponse à cette question que donnent Haspelmath et Rubino est affirmative. Haspelmath considère l'absence de redoublement comme un trait typique des langues SAE (comme le français, l'allemand, le néerlandais...). Dans son étude de la distribution globale du redoublement, Rubino va au-delà de l'hypothèse de Haspelmath. Selon la carte géolinguistique que Rubino a publiée dans le WALS, l'Europe entière représente une aire sans redoublement (exception faite du hongrois). Si son interprétation de la situation européenne est juste, le vieux continent serait l'unique région du monde dans laquelle ce phénomène linguistique ne jouerait aucun rôle.

Dans ma conférence, je me propose de démontrer que, même en Europe, le redoublement total est un phénomène actuel. On peut distinguer deux zones principales : l'une au nord où le redoublement est peu fréquent et s'apparente à la stylistique, et l'autre au sud où le redoublement, plus fréquent, se comporte comme un élément grammatical.

Les conférences du



Note terminologique

Par la suite, exception faite des citations, j'utilise le terme technique „redoublement“ sans différencier entre „réduplication“ ou autres synonymes partiels et „redoublement“.

Situation actuelle

Dès les années 90 nous observons une hausse remarquable d'études linguistiques dédiées à la phénoménologie du redoublement comme stratégie morphologique et/ou morphosyntaxique dans les langues du monde.

Pendant la dernière vingtaine d'années on a traité d'aspects divers de ce phénomène sans résoudre toutes les questions qui se posent, dont une est la distribution géolinguistique du redoublement.

Littérature récente

- Abbi, Anvita. 1992. *Reduplication in South Asian Languages. An Areal, Typological and Historical Study*. New Dehli, etc.: Allied Publishers.
- Ammann, Andreas & Urdze, Aina (éds.). 2007. *Wiederholung, Parallelismus, Redup-ikation. Strategien der multiplen Strukturanwendung*. Bochum: Universitätsverlag Brockmeyer.
- Fabricius, Anne H. 1998. *A Comparative Survey of Reduplication in Australian Languages*. München, Newcastle: LINCOM Europa.
- Hurch, Bernhard (éd.). 2005. *Studies on reduplication*. Berlin, New York: Mouton de Gruyter.
- Inkelas, Sharon & Zoll, Cheryl. 2005. *Reduplication. Doubling in Morphology*. Cambridge: Cambridge University Press.
- Kouwenberg, Silvia (éd.). 2003. *Twice as meaningful. Reduplication in Pidgins, Creoles and other contact languages*. Westminster: Battlebridge.
- Lindström, Jan. 1999. *Vackert, vackert! Syntaktisk reduplikation i svenska*. Helsingfors: Svenska Litteratursällskapet i Finland.
- Michaud, Alexis & Morgenstern, Aliyah (éds.). 2007. *La réduplication*. (= Faits de Langues no. 29). Paris: Ophrys.
- Müller, Hans-Georg. 2004. *Reduplikation im Türkischen. Morphonologische Untersuchungen*. Wiesbaden: Harrassowitz.
- Raimy, Eric. 2000. *The Phonology and Morphology of Reduplication*. Berlin, New York: Mouton de Gruyter.
- Rubino, Carl. 2005a. „Reduplication.“ En: Haspelmath, Martin et al. (éds.), *The World Atlas of Language Structures*. Oxford: Oxford University Press, 114-117.
- Wälchli, Bernhard. 2005. *Co-compounds and natural coordination*. Oxford: Oxford University Press.

Le passé

L'importance du redoublement n'est pas une découverte récente. Les recherches actuelles ont des prédecesseurs vénérables.

Leur histoire commence déjà relativement tôt dans la deuxième moitié du 19^e siècle.

Des textes „classiques“

- Pott, August Friedrich. 1862. *Doppelung (Reduplikation, Geminination) als eines der wichtigsten Bildungsmittel der Sprache beleuchtet aus Sprachen aller Welttheile.* Lemgo, Detmold: Meyer'sche Hofbuchhandlung.
- Brandstetter, Renward. 1917. *Die Reduplikation in den indianischen, indonesischen und indogermanischen Sprachen.* Luzern: Kantonatsschule.
- Gonda, J. 1952. „The functions of word duplication in Indonesian languages.“ *Lingua* 2, 170-197.
- Moravcsik, Edith. 1978. „Reduplicative constructions.“ En: Greenberg, Joseph H. (éd.), *Universals of Human Language. Volume III: Word Structure.* Stanford, Ca.: Stanford University Press, 297-334.
- Bollée, Annegret. 1978. „Reduplikation und Iteration in den romanischen Sprachen.“ *Archiv für das Studium der neueren Sprachen und Literaturen* 215, 318-336.
- Abbi, Anvita. 1980. *Semantic Grammar of Hindi. A Study in Reduplication.* New Dehli: Bahri.
- Skoda, Françoise. 1982. *Le redoublement expressif: un universal linguistique. Analyse du procédé en grec ancien et en d'autres langues.* Paris: SELAF.
- Wierzbicka, Anna. 1986. „Italian reduplication: cross-cultural pragmatics and illocutionary semantics.“ *Linguistics* 24, 287-315.

Pro domo

J'ai découvert le redoublement comme objet scientifique seulement en 2002 par des moyens détournés:

- A l'origine, j'avais l'intention de chercher des traits structuraux qui nous aident à définir l'aire linguistique méditerranée – dont il n'y a pas beaucoup.
- Cependant, il apparaît que le redoublement total est attesté quasiment partout dans les régions au Sud d'Europe – et aussi au-delà.

Mes propres contributions

- Stolz, Thomas. 2002. „Crosscurrents – the Mediterranean region as a potential linguistic area.“ En: Masala, Carlo (éd.), *Der Mittelmeerraum – Brücke oder Grenze?* Baden-Baden: Nomos, 53-74.
- Stolz, Thomas. 2004. „A new Mediterraneanism: word iteration in an areal perspective. A pilot-study.“ *Mediterranean Language Review* 15, 1-47.
- Stolz, Thomas. 2006. „(Wort-)Iteration – (k)eine universelle Konstruktion.“ En: Fischer, Kerstin & Stefanowitsch, Anatol (éds.), *Konstruktionsgrammatik – von der Anwendung zur Theorie*. Tübingen: Stauffenburg, 105-132.
- Stolz, Thomas. 2007a. „Das ist doch keine Reduplikation! Über falsche Beispiele bei der Suche nach richtigen Beispielen.“ En: Ammann/Urdze 2007, 47-80.
- Stolz, Thomas. 2007b. „Re: duplication. Iconic vs. counter-iconic principles (and their areal correlates).“ En: Ramat, Paolo & Roma, Elisa (éds.), *Europe and the Mediterranean as Linguistic Areas*. Amsterdam, Philadelphia: Benjamins, 317-350.
- Stolz, Thomas. 2008a. „GRAMMATIKALISIERUNG EX NIHILO. Totale Reduplikation – ein potentielles Universale und sein Verhältnis zur Grammatikalisierung.“ En: Stolz, Thomas (éd.), *Grammatikalisierung und grammatische Kategorien*. Bochum: Universitätsverlag Brockmeyer, 83-110.
- Stolz, Thomas. 2008b. „Total reduplication vs. echo-word formation in language contact situations.“ En: Siemund, Peter & Kintana, Noemi (éds.), *Language Contact and Contact Languages*. Amsterdam, Philadelphia: Benjamins, 107-132.

A paraître

- Stolz, Thomas; Stroh, Cornelia & Urdze, Aina. 2010. *Total Reduplication – the areal linguistics of a putative universal.* Amsterdam, Philadelphia: Benjamins. 650 pp.

[Cette monographie présente les résultats d'un projet de typologie linguistique (financé par la DFG) que nous avons conduit dès 2004 jusqu'à 2009. Ma conférence d'aujourd'hui se réfère à un chapitre de ce livre à paraître.]

Définition préliminaire

Le redoublement total, qu'est-ce que c'est?

Une définition (un peu compliquée et insuffisante):

Le redoublement total est une construction linguistique qui requiert qu'une séquence d'éléments phoniques qui forment un mot ou un morphème lexical (c.-à-d. qui n'est pas vide en ce qui concerne la sémantique) soit répétée immédiatement après sa première occurrence dans le même constituant syntactique afin de former une nouvelle unité fonctionnelle qui a une sémantique plus ou moins distincte de celle du mot isolé.

Un exemple africaine

- Souahili (Brauner/Herms 1986: 273)

Kuna n-chi n-dogo n-dogo zi-li-zo maskini

Exi CI5-pays CI5-petit CI5-petit CI5.PI-Dem-CI5.PI pauvre

‘Il y a des petits pays qui sont pauvres.’

‘There are small countries which are poor.’

[L’adjectif *-dogo* ‘petit’ est attesté deux fois comme attribut du nom *-chi* ‘pays’ dans la même phrase nominale. Le redoublement de l’adjectif a la fonction d’exprimer le pluriel parce que les préfixes des noms et adjectifs appartenant à la classe 5 du Souahili ne distinguent pas entre les deux catégories de nombres. Donc la construction à redoublement //*ndogo ndogo*// n’est pas synonyme avec l’adjectif simple //*ndogo*//. Je souligne que le redoublement d’adjectif n’intensifie pas la sémantique de l’adjectif!]

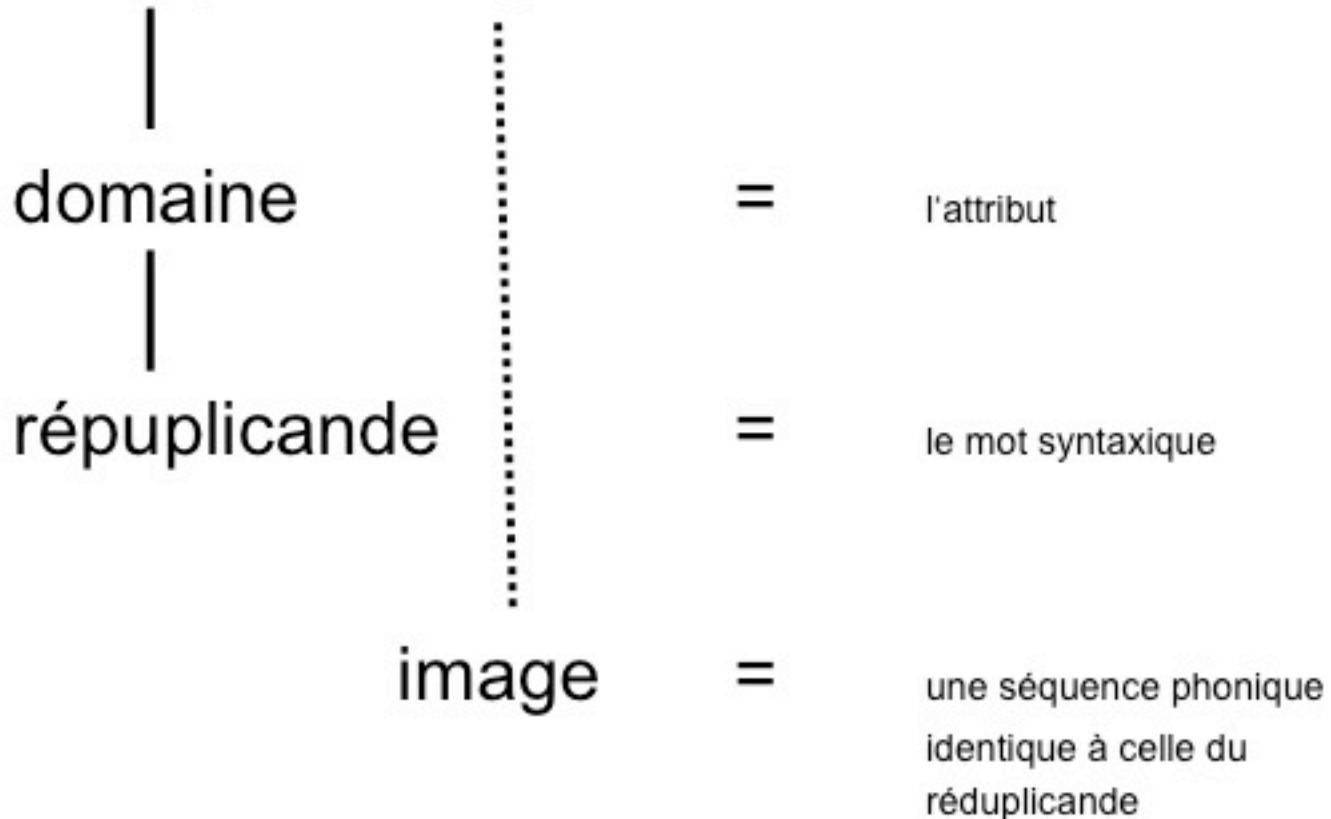
Modèle

En ce qui concerne la terminologie analytique je me réfère à Mel'čuk (1996: 41) qui distingue trois catégories essentielles dans son modèle du redoublement:

- Domaine d'une réduplication:
 - La partie du signifiant d'une signe soumise à une réduplication est appelée le domaine de cette réduplication.
- Réduplicande d'une réduplication
 - La partie du domaine d'une réduplication actuellement affectée par elle (c'est-à-dire recopiée) est appelée le réduplicande.
- Image d'une réduplication
 - La copie du réduplicande produite par une réduplication est appelée l'image du réduplicande (ou de la réduplication).

Souahili

n-chi n-dogo n-dogo



Un problème élémentaire

- Il est souvent très difficile de ne pas confondre le redoublement total, qui a des fonctions strictement grammaticales ou lexicales, avec la répétition purement stylistique et pragmatique, qui exprime l'emphase et/ou l'insistance de l'énonciateur.
 - Ce problème est évident dans des contextes où emphase et intensification semblent coïncider.

Répétition ou redoublement?

Français [LPP French, 45]

*Les grandes personnes sont
décidément **[très très]** bizarres.*

très très = « plus de très » ?

Je ne parlerai pas de:

- Redoublement partiel (y compris procès phonologiques comme la gémination consonantique)
- Redoublement total à variation phonique
- Redoublement à distance
- Redoublement lexical (y compris les co-composés)
- Construction corrélatives
- Négation multiples et constructions similaires
- Concordance grammaticale
- Triplication (et d'autres procédés plus étendus)
- Parallélisme stylistique
- Etc. etc. etc.

Encadrement

Mon étude:

- est synchronique
- se base sur les principes de la typologie fonctionnelle
- a des buts géolinguistiques
- se sert des méthodes quantitatives de la statistique linguistique
- contribue à l'Euro-linguistique

Un trait universel(?)

- Dans la littérature récente, on cite très souvent Moravcsik (1978: 328 note 13) qui a formulé encore prudemment:

"[w]hereas I have indeed found many examples of languages that have both partial and total reduplication [...] and also some that do have total reduplication but may not have partial [...], I am uncertain if there are any languages at all without any kind of reduplication. [...] Should [...] some [...] meaning category be universally expressible through total reduplication, then the proposed typological statement about the dependence of partial reduplication on total reduplication in a language should be replaced by an unrestricted universal statement asserting the universal occurrence of total reduplication."

Un trait universel – sans doute! Ou quoi?

- La prudence de Moravcsik a disparu successivement de sorte que Raimy (2000: 97) et d'autres experts contemporains du redoublement se réfèrent (normalement en passant) au texte classique de Moravcsik en généralisant: eux, ils acceptent l'idée que le redoublement total « **may be a true language universal** ».
- Puisque l'intérêt des linguistes se concentre sur le redoublement partiel et ses implications pour la théorie de la morphologie linguistique, l'universalité supposée du redoublement total n'a jamais été vérifié par l'empirie.
- Le premier essai empirique modern est celui de Rubino (2005a) – en combinaison avec un deuxième texte à peu près identique (Rubino 2005b).
- Pour lui, à proprement parler, **le redoublement total n'est pas un trait universel des langues humaines** quoiqu'il soit très répandu en perspective mondiale. Pourtant il y a des lacunes dans sa distribution géolinguistique – et une de ces lacunes se trouve en Europe.

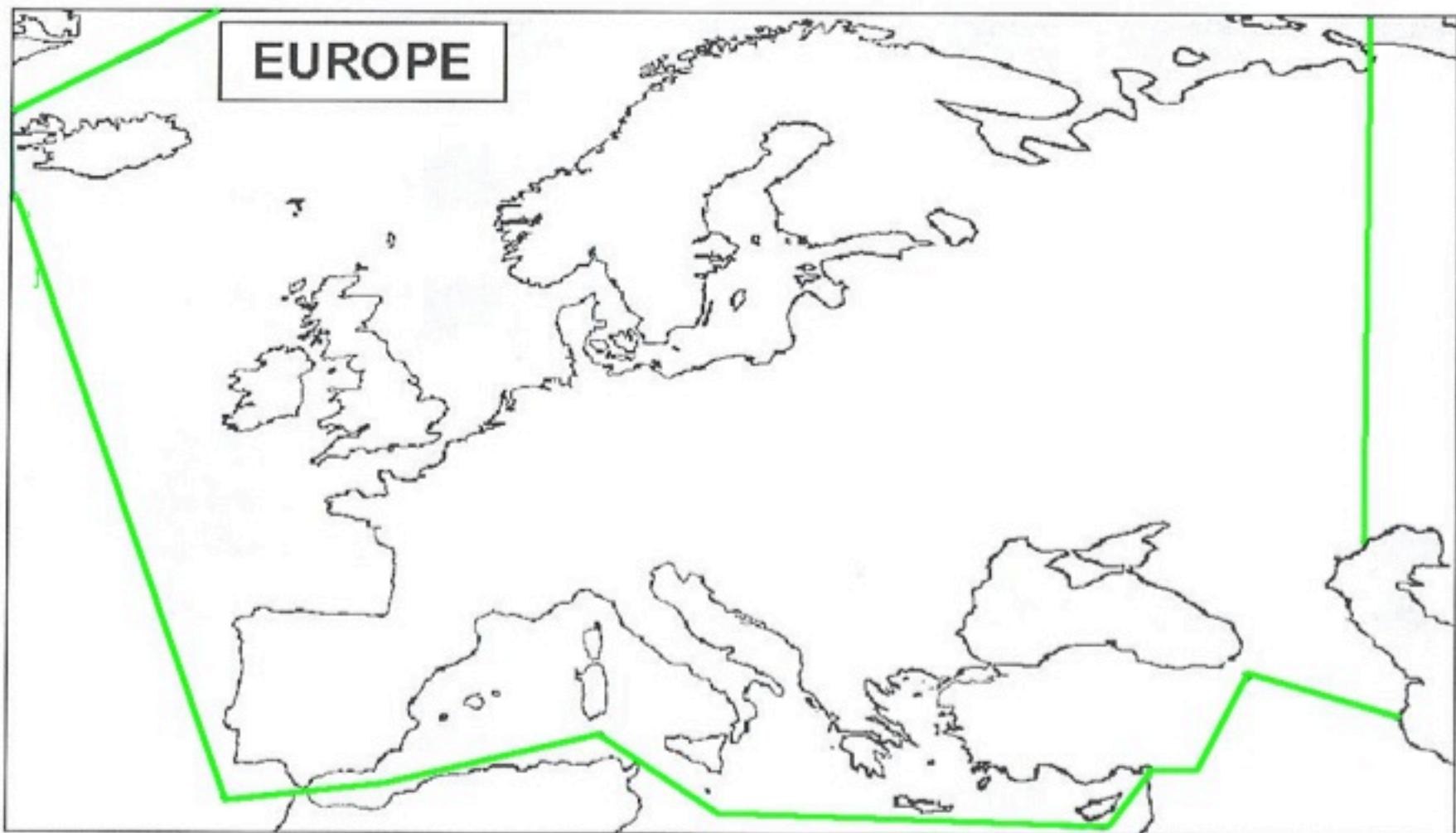
L'Europe – un continent sans redoublement?

- Il y a deux hypothèses géolinguistique récentes qui suggèrent qu'en Europe, le redoublement total est un phénomène inconnu:
 - Selon Haspelmath (2001), les langues européennes du type SAE (c.-à-d. les langues de l'Europe ouest-central et des régions limitrophes) ne connaissent aucun procédé réduplicatif systématique.
 - [Cette hypothèse n'exclut pas l'existence du redoublement dans d'autres langues européennes (plus périphériques).]
 - Rubino (2005a-b) défend l'idée que l'Europe, dans sa quasi-totalité, est une aire linguistique dépourvue des mécanismes réduplicatifs productifs.
 - [Cette hypothèse est plus radical parce qu'elle comprend aussi la majorité des langues non-SAE du continent européen. Elle définit l'Europe comme une des régions linguistiques exceptionnelles du monde.]

La distribution global du redoublement selon Rubino (2005a)



L'Europe – qu'est-ce que c'est?



Répartition

- Rubino (2005a) a contrôlé 367 langues qui se divisent en trois groupes:
 - Langues à redoublement total et partiel
 - 277 ~ 75%
 - Langues exclusivement à redoublement total
 - 34 ~ 9%
 - Langues sans aucun procédé réduplicatif
 - 56 ~ 15%

La situation en Europe selon Rubino (2005a)

- 6 langues (= 26%) à redoublement productif:
 - abkhaz, arménien (oriental et occidental), géorgien, hongrois, hunzib, turque
- 17 langues (= 74%) sans redoublement productif:
 - allemand, anglais, **basque**, *catalan*, espagnol, français, écossais, finnois, **gallois**, **grec**, irlandais, **italien**, **lezguien**, **mari (patûre)**, *portugais*, russe
 - Les caractères gras identifient les langues qui ne sont pas classifiées correctement par Rubino (2005a). Les glossonymes en italiques indiquent les langues dont la classification est au moins douteuse.

Sources d'information = sources d'erreurs

- Au moins une bonne partie des classifications fautives résulte des grammaires que Rubino (2005a) a consulté parce qu'elles passent le redoublement sous silence.
- C'est le cas p.ex. avec:
 - Le basque:
 - La grammaire descriptive écrite en anglais par Saltarelli at al. (1988) ne dit rien du redoublement – et elle est la source d'information de Rubino (2005a),
 - qui ne connaît pas la grammaire écrite en français par Lafitte (1979) qui montre clairement que le basque emploie le redoublement très frequemment.

Composition du cadre des langues

- Le choix et le nombre des langues sont deux facteurs qui ont une influence décisive sur l'idée de l'Europe comme le continent où le redoublement est un phénomène marginal. P.ex.
 - avec les langues turques et les langues fennou-ougriennes de l'Eurasie, Rubino (2005a) exclue deux groupes importants de langues à redoublement grammatical,
 - leur inclusion dans le cadre des langues aurait produit un image très différent de la distribution géolinguistique du redoublement en Europe.

Productivité

- Rubino (2005a: 115) parle de la présence ou absence de procédés productifs comme critère essentiel de sa pratique classificatoire.
- Malheureusement, il n'explique pas comment il détermine la productivité.
- C'est pourquoi il ne nous donne aucune valeur indicative pour calculer la productivité.

Deux études quantitatives

- A fin de déterminer le rôle du redoublement en Europe, je vais présenter les résultats de deux études quantitatives.
- Ces études se basent sur deux corpus littéraires parallèles – à savoir:
 - L'original français et les traductions de « Le Petit Prince » (112 langues européennes et 16 langues non-européennes). Ce texte contient **1.650 phrases**.
 - L'original anglais et les traductions de « Harry Potter, tome I » (41 langues d'Europe et 6 langues situées hors de l'Europe). Ce texte contient **11.270 phrases**.

Je clarifie

- Dans ce qui suit je parle tout simplement de **langues** –
 - Ce terme comprend
 - variétés standard
 - variétés non-standard
 - langues officielles
 - langues régionales
 - dialectes

(C6) *Le Petit Prince* corpus: sample languages

INDO-EUROPEAN	96
Romance: Aragonese, Aranese, Aromanian, Asturian, Badiota, Bergamasco, Bolognese, Catalan, Corsican, Extremenio, Franco-Provençal (France), Franco-Provençal (Val d'Aosta), French, Friulan, Galego, Gascon, Gherdëina, Italian, Languedocien, Milanese, Moldavian, Napolitan, Niçard, Piemontese, Provençal (France), Provençal (Italy), Rumantsch Grischun, Surmeir, Surselvian, Rumanian, Sardinian, Valladér, Venetian, Wallon	34
Germanic: Allemannic (Baden), Alsatian, Alvdalsk, Austrian (Upper), Bärndüütsch, Berlinian, Carinthian, Danish, Drents, Dutch, English, Faroese, Franconian, Frisian (Saterland), Frisian (West), German, Hassian, Icelandic, Kölsch, Limburgian (Northern), Limburgian (Southern), Lothringian, Low German (East Friesland), Low German (East Holstein), Luxembourgish, Norwegian (Bokmål), Palatinate, Plautdietsch, Saarlandish, Swabian, Swedish, Tyrolean (Northern), Tyrolean (Southern), Yiddish	34
Slavic: Bielorussian, Bosnian, Bulgarian, Croatian (Burgenland), Croatian (Croatia), Czech, Macedonian, Molisano, Polish, Russian, Serbian, Slovak, Slovenian, Sorbian (Upper), Ukrainian	15
Celtic: Breton, Irish, Scots-Gaelic, Welsh	4
Baltic: Latvian, Lithuanian	2
Indo-Aryan: Kurmanci, Lovari, Zazaki	3
Other: Albanian (Gheg), Albanian (Tosk), Armenian (East), Greek	4
NON-INDO-EUROPEAN	16
Turkic: Kazakh, Tatar, Turkish	3
Uralic: Estonian, Finnish, Hungarian, Saami (Inari), Saami (Northern), Saami (Skolt), Udmurt	7
Caucasian: Abkhaz, Chechen, Georgian	3
Afro-Asiatic: Aramaic (East), Maltese	2
Other: Basque	1
TOTAL: INDO-EUROPEAN + NON-INDO-EUROPEAN	112

Avant de compter

Les cinq exemples suivants vous donnent une idée du type de constructions qui passent pour représentantes du redoublement.

Exemples européens I

(1) Sarde [LPP Sardinian VII.42]

Connosco	un'	astru	inue	b'	at
connaître:1Sg	un	étoile	où	là	avoir.3Sg
unu	Sennore	ruju	ruju		
un	monsieur	rouge:M	rouge:M		

‘Je connais une planète où il y a un monsieur (**tout**)
cramoisi’.

Exemples européens II

- (2) Albanais (Tosk) [LPP Tosk XXVI.174]

*Dhe ra **butë-butë** ashtu sikundër bië një pemë*
et tomber:Aor.3Sg douce-douce si comme tombe un arbre
'Il tomba (très) doucement comme tombe un arbre.'

Exemples européens III

(3) Grec [LPP Greek XIV.8]

Sto **kátō-kátō** é douleiá tou éxei
dans:Def **dessous-dessous** Def.F travail Por.3Sg avoir:3Sg
kai kápoio nóēma
et certain sens

‘Au moins son travail a-t-il un sens.’

Exemples européens IV

(4) Basque [LPP Basque II.16]

<i>Neri</i>	<i>begira</i>	<i>serio-serio</i> ,	<i>gizontto</i>	<i>harrigarri</i>	<i>bat</i>
1Sg:Dat	regarde	grave-grave	homm:Dim	remarcable	un
<i>nedukan</i>	<i>begien</i>		<i>aurrean.</i>		
1Sg:avoir		œil:Gen	front:Iness		

‘Et j’ai vu un petit bonhomme tout à fait extraordinaire qui me considérait (**très**) gravement.’

Exemples européens V

(5) Gallois [HP I Welsh 15]

Yr unig beth yr hoffai Harri ynghylch ei olwg

Def unique chose Rel aimer:Pret.3Sg Harry environ Por.3 air

*ei hun oedd y graith **fain** **fain***

Por.3 même être:Imperf.3 Det cicatrice mince mince

ar ffurf mellten ar ei dalcen

à forme éclair à Por.3 front

‘La seule chose que Harry aimait bien dans son apparence physique, c’était la (**très**) **fine** cicatrice qu’il portait sur le front et qui avait la forme d’un éclair.’

Étude I: Tokens en Europe

- La statistique des tokens du redoublement dans les versions de « Le Petit Prince » montre que
 - le phénomène est absent dans 36 langues dont la moitié sont langues germaniques,
 - le phénomène est relativement fort dans des langues qui ne n'appartiennent pas à la famille indo-européenne ou sont langues romanes régionales,
 - le redoublement est plus fort dans le Sud et l'Est du continent.

(C7) TR in the *Le Petit Prince*-corpus I: tokens per language

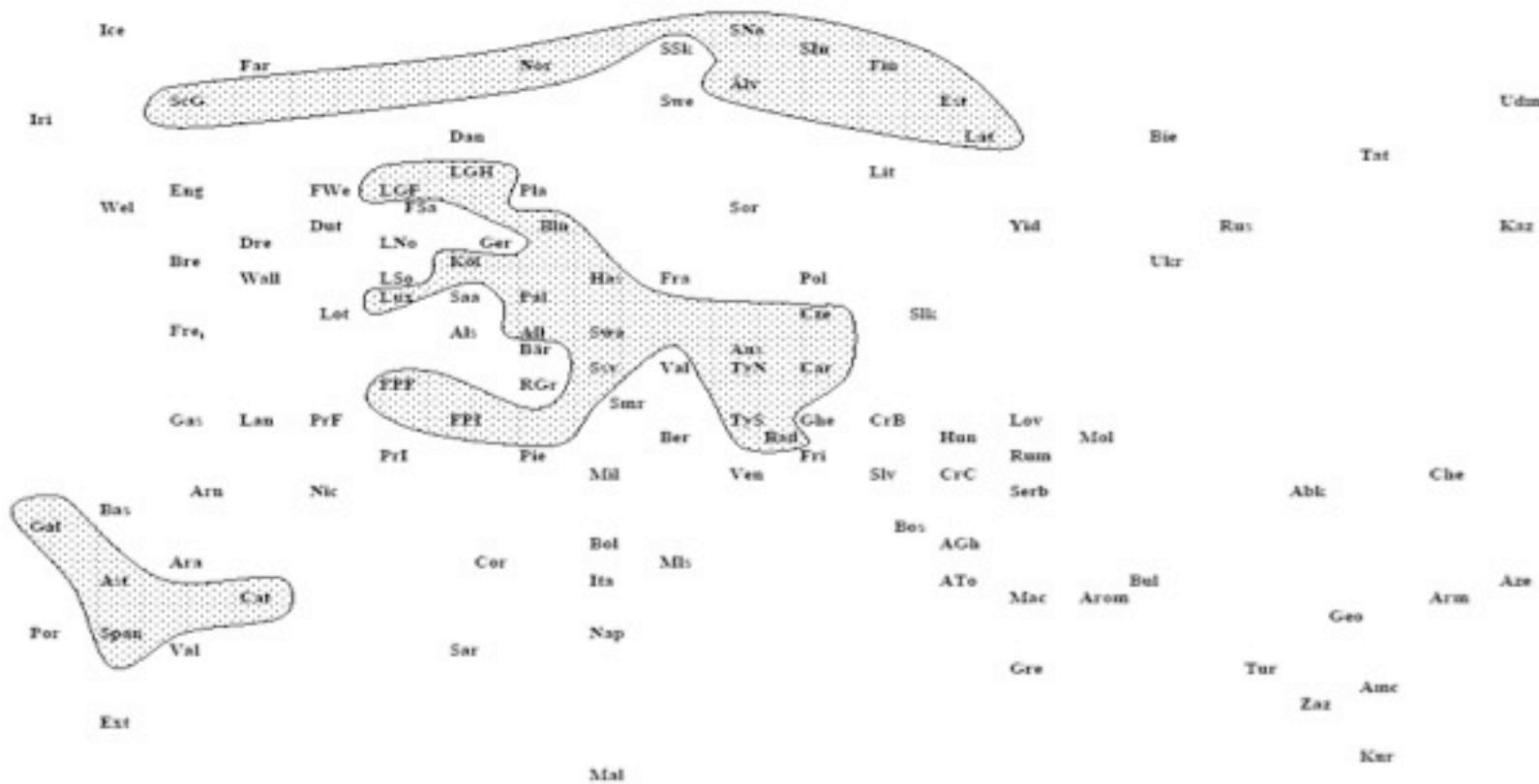
Rank	Language	Tokens
1	Basque	44
2	Azeri	36
3	<i>Napoletan</i>	35
4	Tatar	25
5-6	Chechen; Corsican	24
7	Maltese	23
8	Abkhaz	21
9	Turkish	19
10-12	<i>Niçard; Sardinian; Udmurt</i>	18
13	<i>Provençal (Italy)</i>	17
14-16	Albanian (Tosk); Aramaic (East); Armenian (East)	13
17-19	Greek; Hungarian; Venetian	11
20	<i>Aranese</i>	9
21	Molisano	8
22-24	<i>Aragonese; Bolognese; Welsh</i>	7
25-28	Georgian; Italian; Milanese; Zazaki	6
29-30	<i>Bielorusian; Valenciano</i>	5
31-34	<i>Aromunian; Bergamasco; Kurmanci; Piemontese</i>	4
35-39	Bosnian; Gascon; Kazakh; Portuguese; Yiddish	3
40-51	<i>Albanian (Gheg); Croatian (Burgenland); Danish; Extremoño; Faroese; Frisian (Saterland); Frlulan; Languedocian; Provençal (France); Rumantan; Rumantsch Grischun; Russian</i>	2
52-86	<i>Alsatian; Bärndüütsch; Breton; Bulgarian; Croatian (Croatia); Dreents; Dutch; English; Franconian; French; Frisian (West); German; Gherdëina; Icelandic; Irish; Limburgian (Northern); Limburgian (Southern); Lithuanian; Lothringian; Lovari; Macedonian; Moldavian; Plautdietsch; Polish; Saami (Skolt); Saarlandish; Serbian; Slovak; Slovenian; Sorbian (Upper); Surmeir; Swedish; Ukrainian; Vallader; Wallon</i>	1
87-112	<i>Allemanic (Baden); Alvdalsk; Asturian; Austrian (Upper); Badtota; Berlinian; Carinthian; Catalan; Czech; Estonian; Finnish; Franco-Provençal (France); Franco-Provençal (Val d'Aosta); Galego; Hassian; Kölsch; Latvian; Low German (East Friesland); Low German (East Holstein); Luxemburgish; Norwegian (Bokmål); Palatinate; Saami (Inari); Saami (Northern); Scots Gaelic; Spanish; Sursilvan; Swabian; Tyrolean (Northern); Tyrolean (Southern)</i>	0

Cartographie I: peu de tokens

- Les quatre cartes suivantes démontrent que
 - la tendance à éviter des constructions à redoublement suit une logique géolinguistique parce que
 - les langues à aversion pour le redoublement se concentrent dans une aire rectangulaire qui occupe la partie centrale-nord-ouest de l'Europe (avec des petits îlots hors de ce territoire).

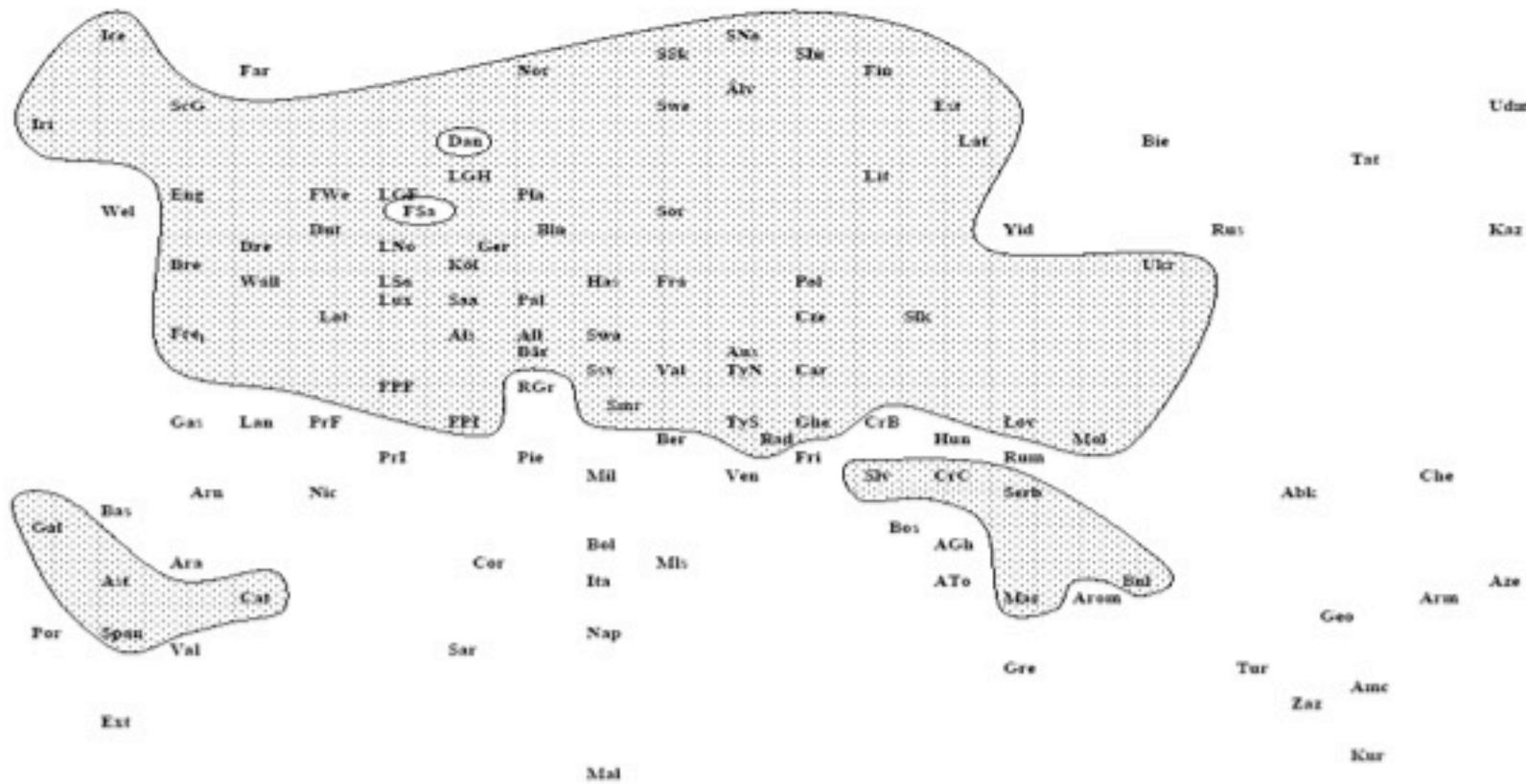
Tokens: Langues à fréquence zéro

(CS) Token frequency n = 0



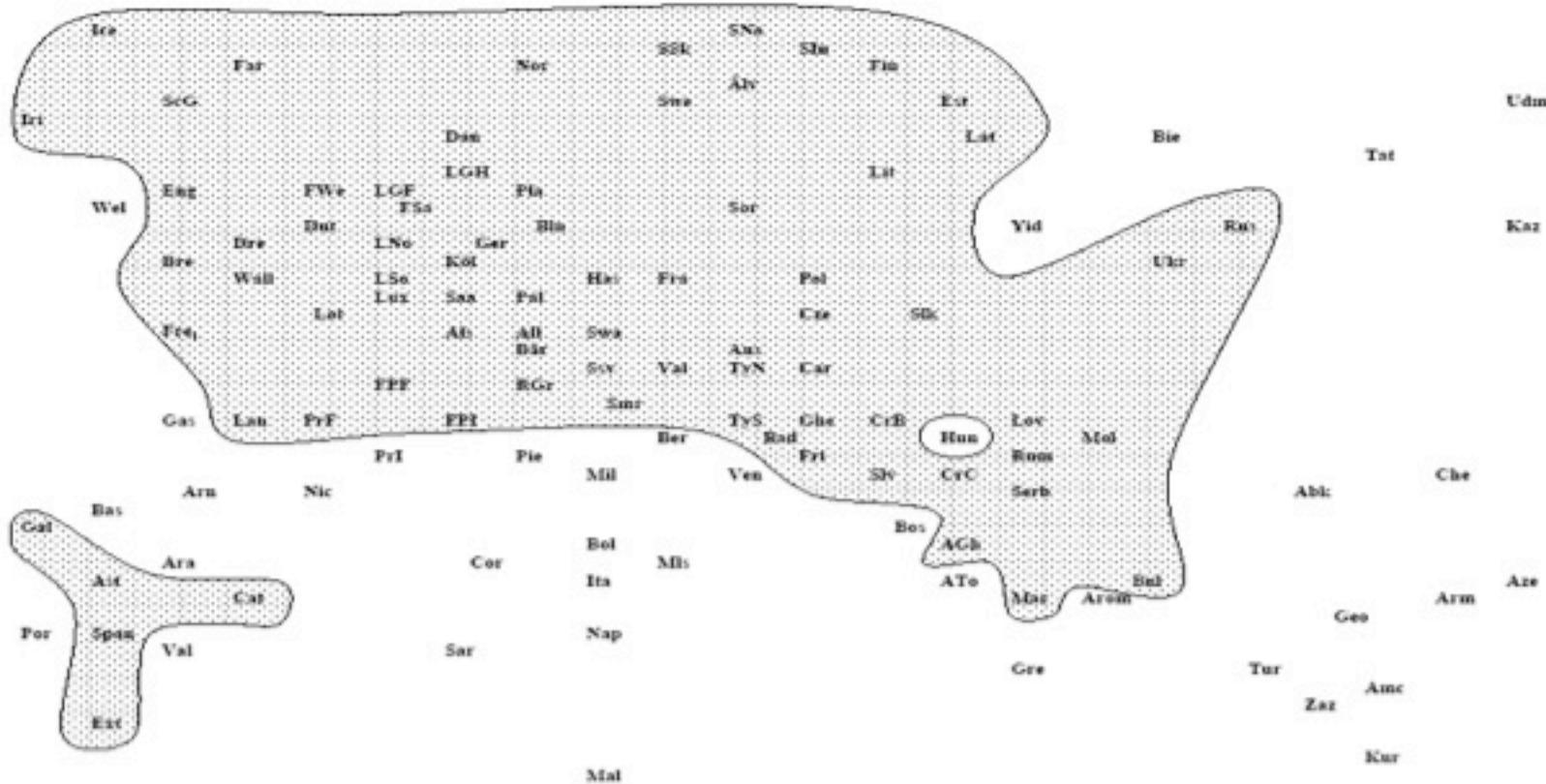
Tokens: Langues à fréquence minime

(C9) Token frequency $n \leq 1$.



Tokens: Langues à fréquence n ≤ 2

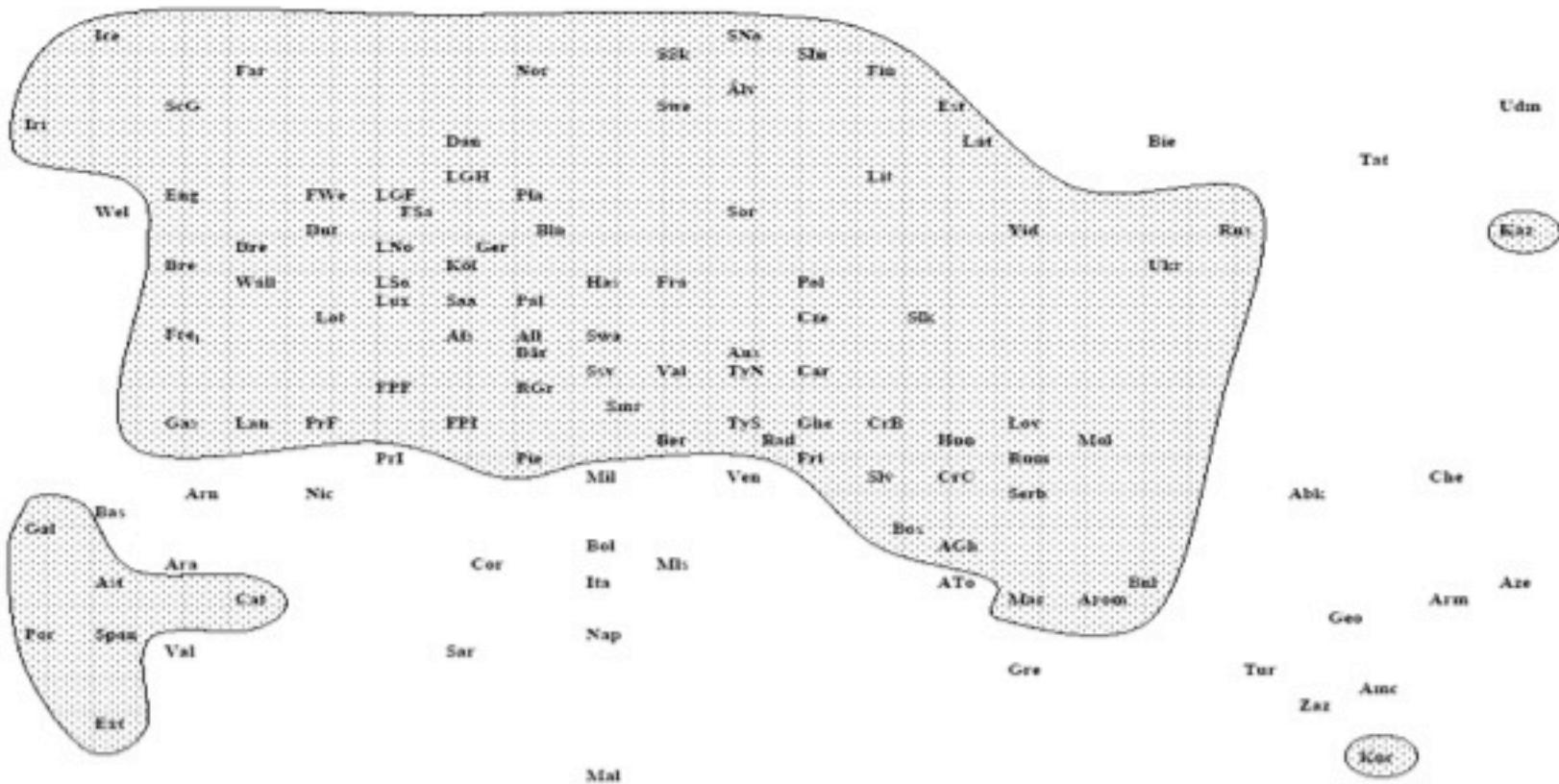
(C10) Token frequency n ≤ 2



Tokens:

Langues à fréquence $n \leq 4$

(C11) Token frequency $n \leq 4$



Cartographie I: Beaucoup de tokens

- Les 6 cartes suivantes servent comme contre-épreuve. Elles démontrent que
 - les langues qui emploient le redoublement relativement souvent se trouvent en majorité dans les régions méridionales et orientales du continent,
 - elles ne se superposent pas avec les langues à aversion pour le redoublement.

Tokens:

Langues à fréquence $n \geq 17$

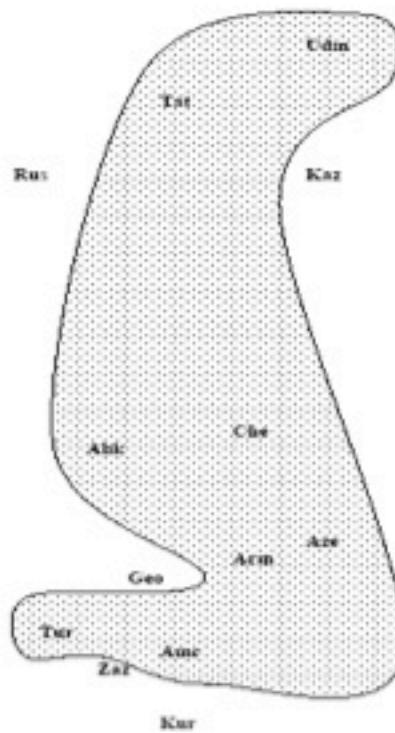
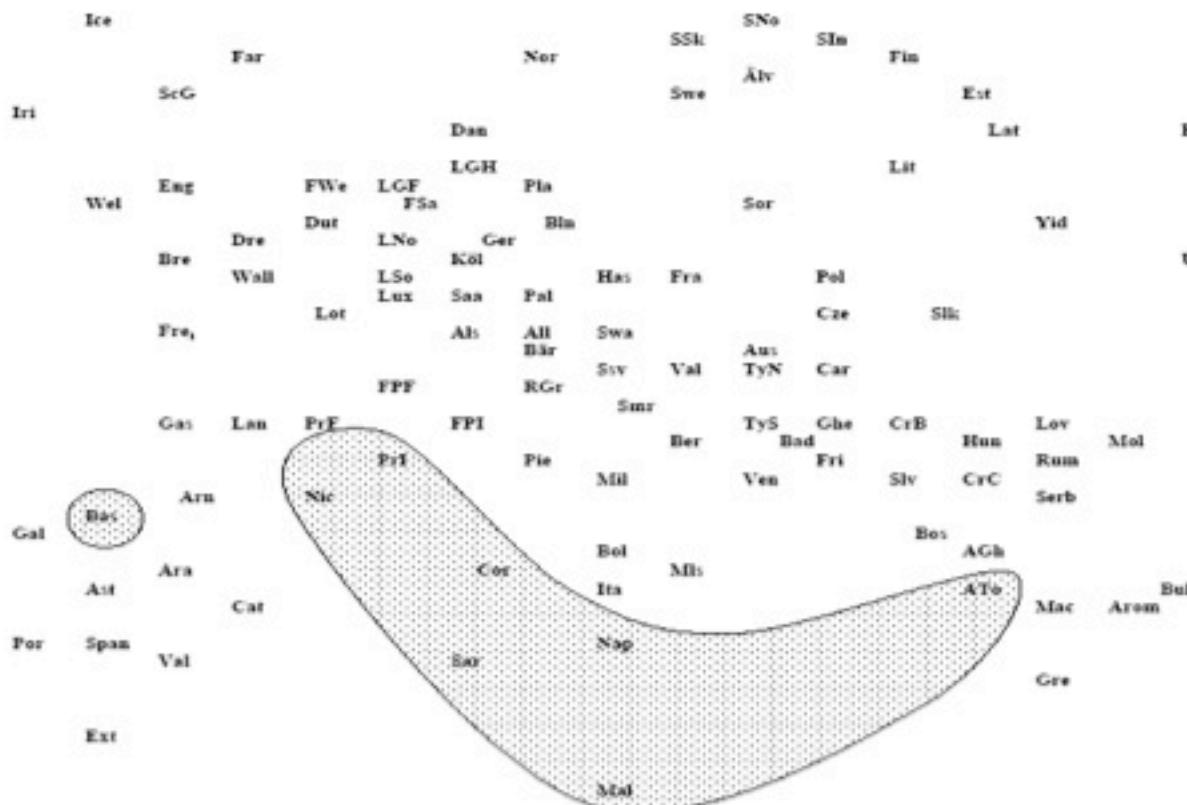
(C12) Token frequency $n \geq 17$



Tokens:

Langues à fréquence $n \geq 13$

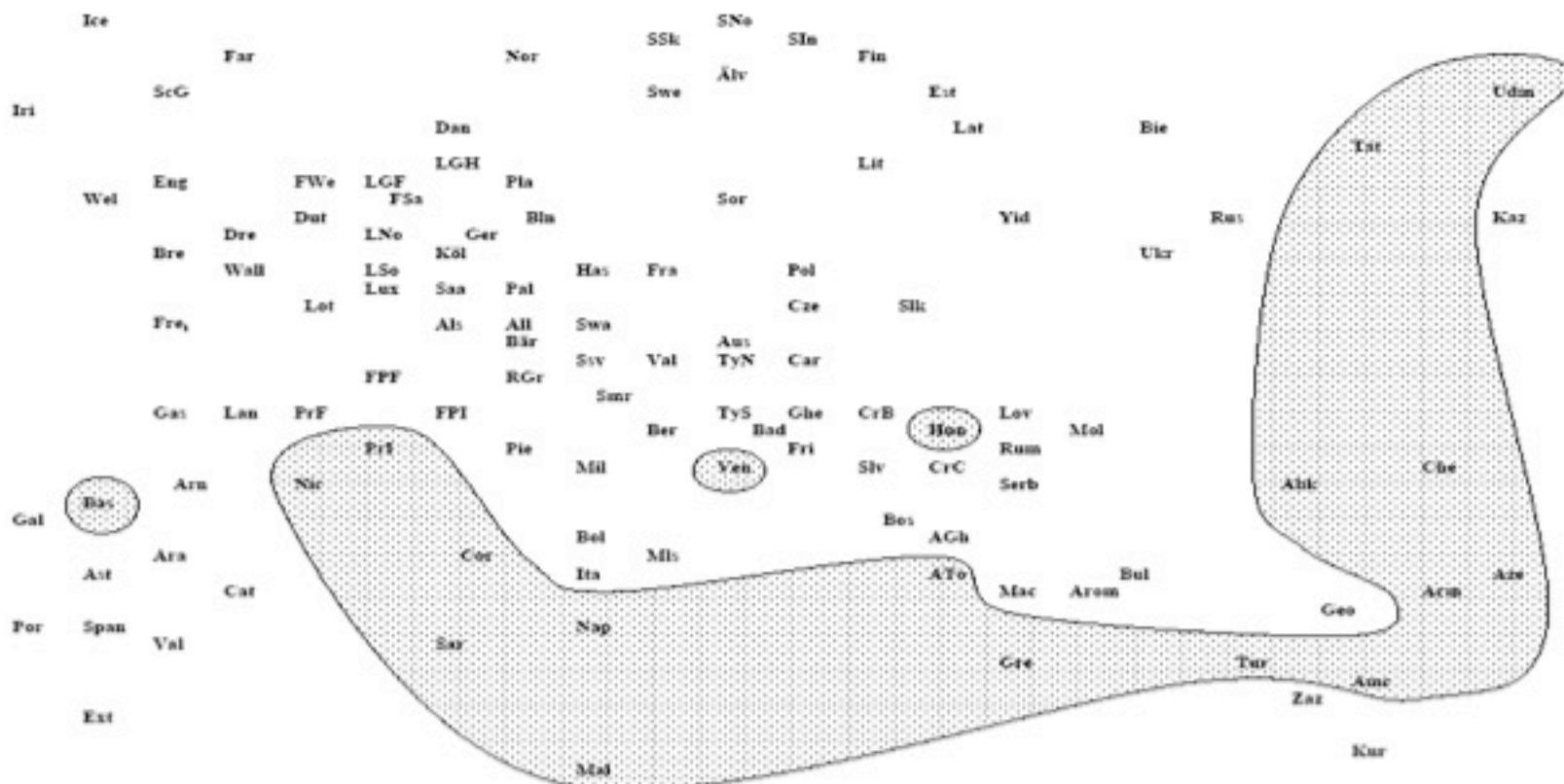
(C13) Token frequency $n \geq 13$



Tokens:

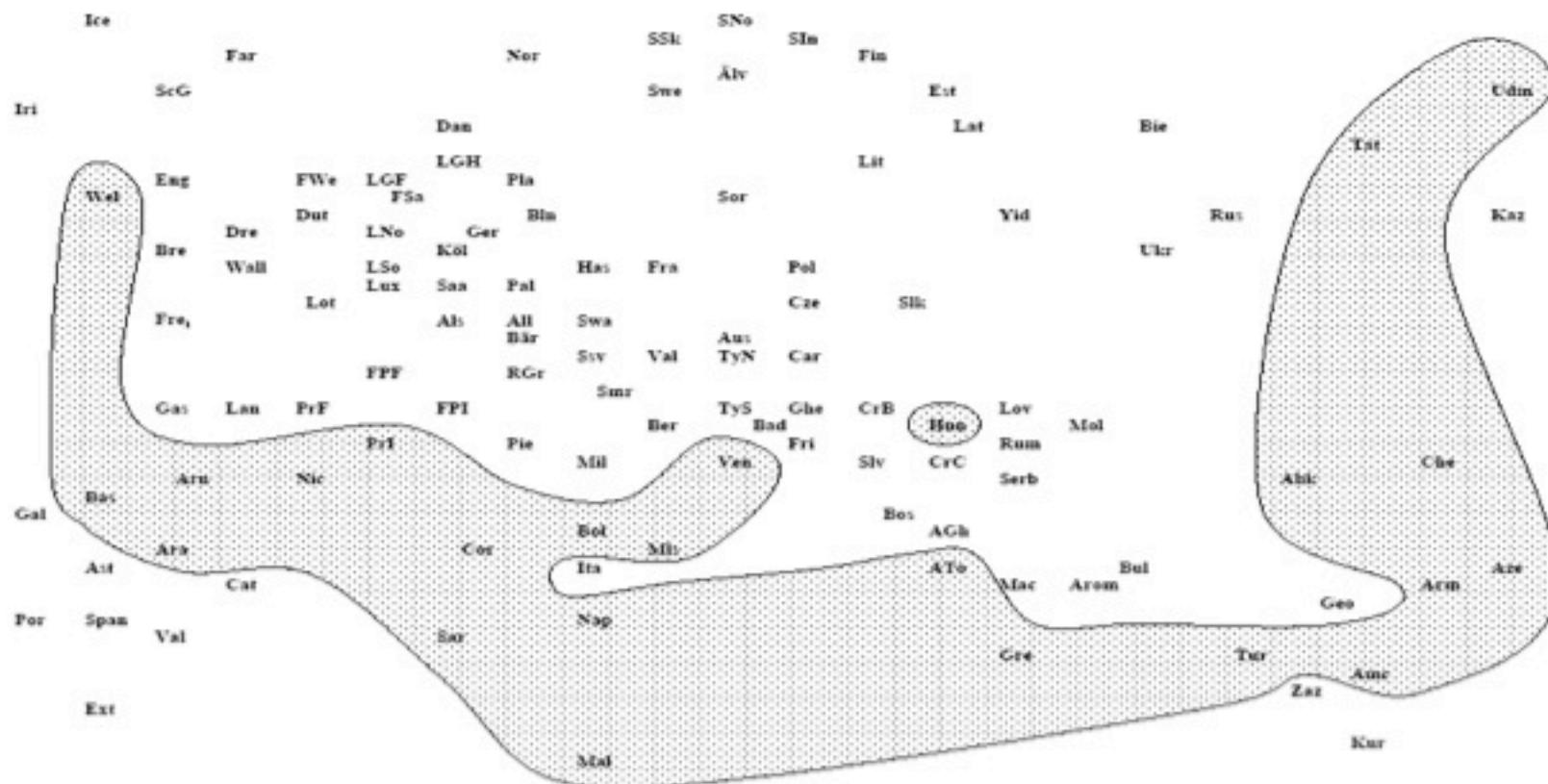
Langues à fréquence $n \geq 11$

(C14) Token frequency $n \geq 11$



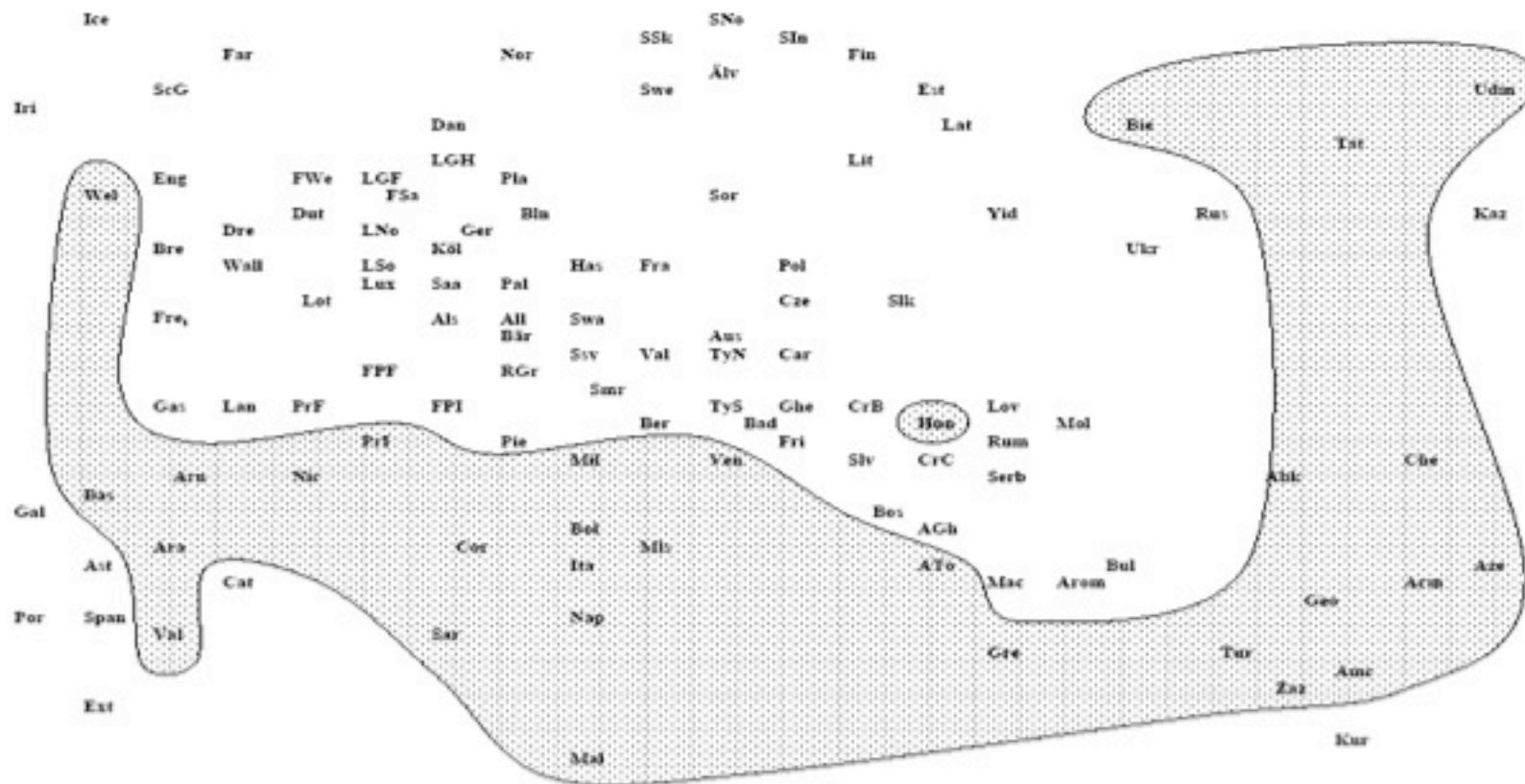
Tokens: Langues à fréquence $n \geq 7$

(C15) Token frequency $n \geq 7$



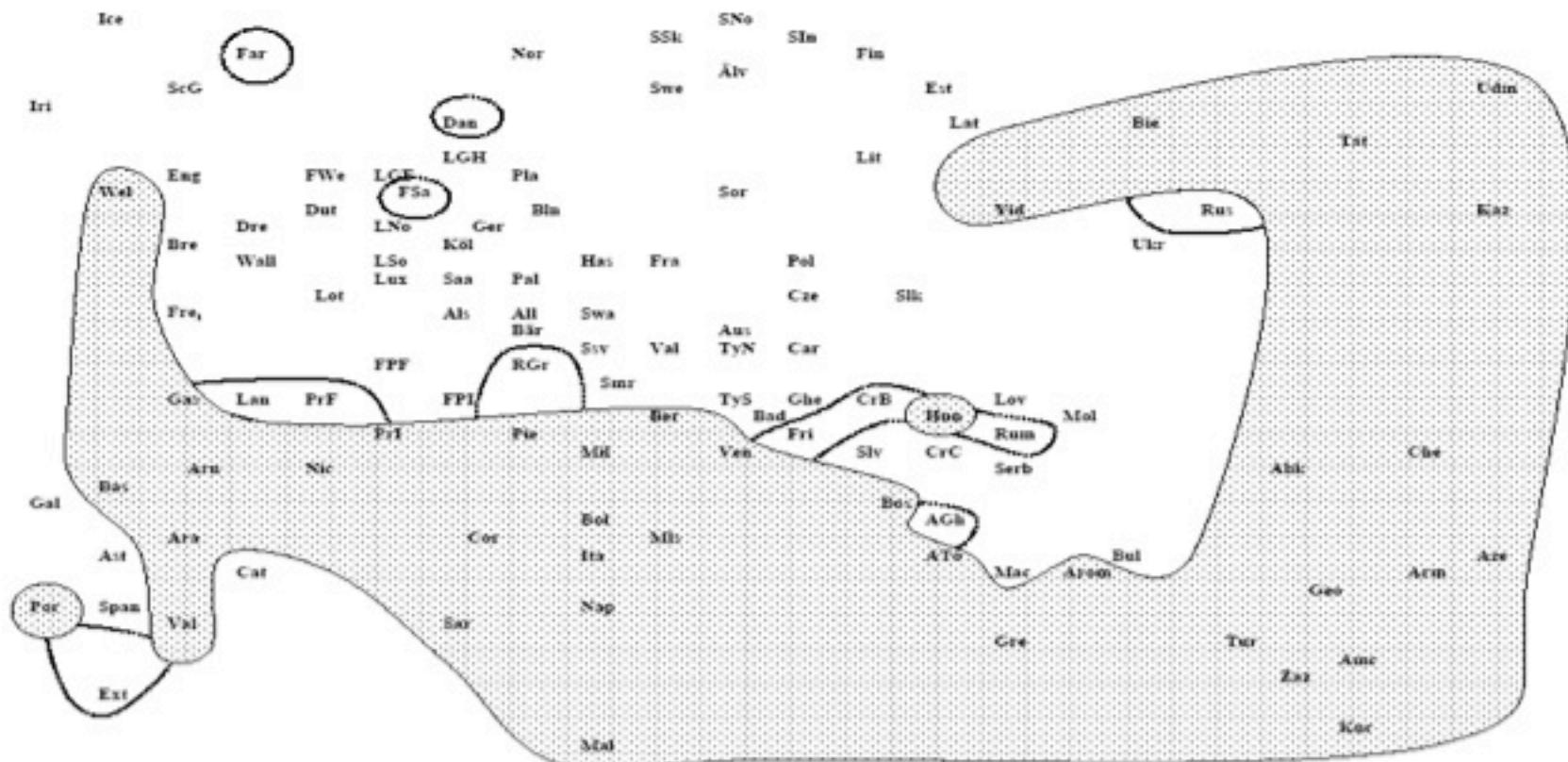
Tokens: Langues à fréquence $n \geq 5$

(C16) Token frequency $n \geq 5$



Tokens: Langues à fréquence $n \geq 3$ et $n \geq 2$

(C17) Token frequency $n \geq 3$ and $n \geq 2$



Études I: Types en Europe

- La statistique des types du redoublement dans les versions de « Le Petit Prince » confirme les tendances mentionnées ci-dessus:
 - les langues germaniques limitent le nombre des lexèmes qui subissent le redoublement à une quantité négligeable,
 - tandis que les langues romanes régionales et les langues non-indo-européennes permettent à relativement beaucoup de lexèmes d'être redoublés.
 - Encore une fois, la fréquence du phénomène est forte dans le Sud et l'Est d'Europe.

(C18) TR in the *Le Petit Prince*-corpus II: types per language

Rank	Language	Types
1	Basque	29
2	Azeri	27
3	Tatar	19
4	<i>Napoletan</i>	17
5	Chechen	16
6	Abkhaz	15
7	Turkish	13
8-9	<i>Armenian (East); Udmurt</i>	10
10-11	<i>Corsican; Sardinian</i>	9
12	Hungarian	8
13-14	<i>Greek; Maltese</i>	7
15-17	<i>Albanian (Tosk); Milanese; Venetian</i>	6
18-21	<i>Aragonese; Bielorusian; Welsh; Zazaki</i>	5
22-24	<i>Aramaic (East); Aromanian; Italian</i>	4
25-33	<i>Bergamasco; Bolognese; Gascon; Kazakh; Kurmanci; Niçard; Piemontese; Portuguese; Provençal (Italy)</i>	3
34-46	<i>Albanian (Gheg); Aranese; Danish; Faroese; Friulan; Georgian; Languedocien; Molisano; Provençal (France); Rumanian; Rumantch Grischur; Valenciano; Yiddish</i>	2
47-86	<i>Alsatian; Bärndöitsch; Bosnian; Breton; Bulgarian; Croatian (Burgenland); Croatian (Croatia); Drents; Dutch; English; Extremello Franconian; French; Frisian (Saterland); Frisian (West); German; Gherdëina; Icelandic; Irish; Limburgian (Northern); Limburgian (Southern); Lithuanian; Lotharingian; Lovari; Macedonian; Moldavian; Plautdietsch; Polish; Russian; Saami (Skolt); Saarlandish; Serbian; Slovak; Slovenian; Sorbian (Upper); Surmeir; Swedish; Ukrainian; Vallader; Wallon</i>	1

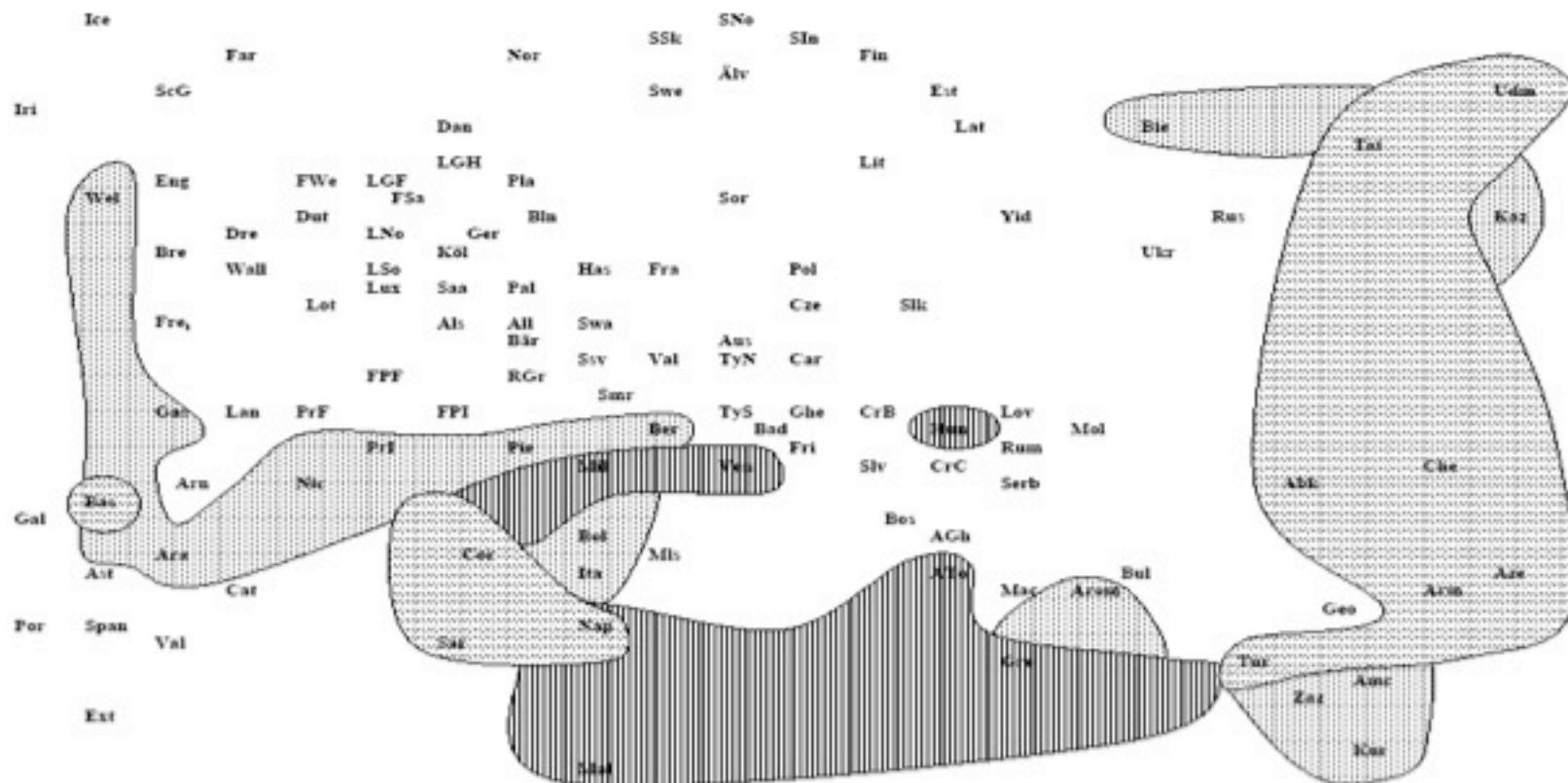
Cartographie II: Types en Europe

- Les deux cartes suivantes confrontent la distribution géolinguistique en Europe des langues à préférence pour le redoublement et celle des langues à fréquence minime du phénomène.
 - Cette confrontation montre qu'il y a une dichotomie nette entre
 - Langues à préférence pour le redoublement que se situent dans un demi-cercle méridional-oriental (avec une extension occidentale) et
 - Langues à aversion pour le redoublement au centre – quasiment encerclées par les membres du premier groupe de langues.

Types: Langues à fréquence

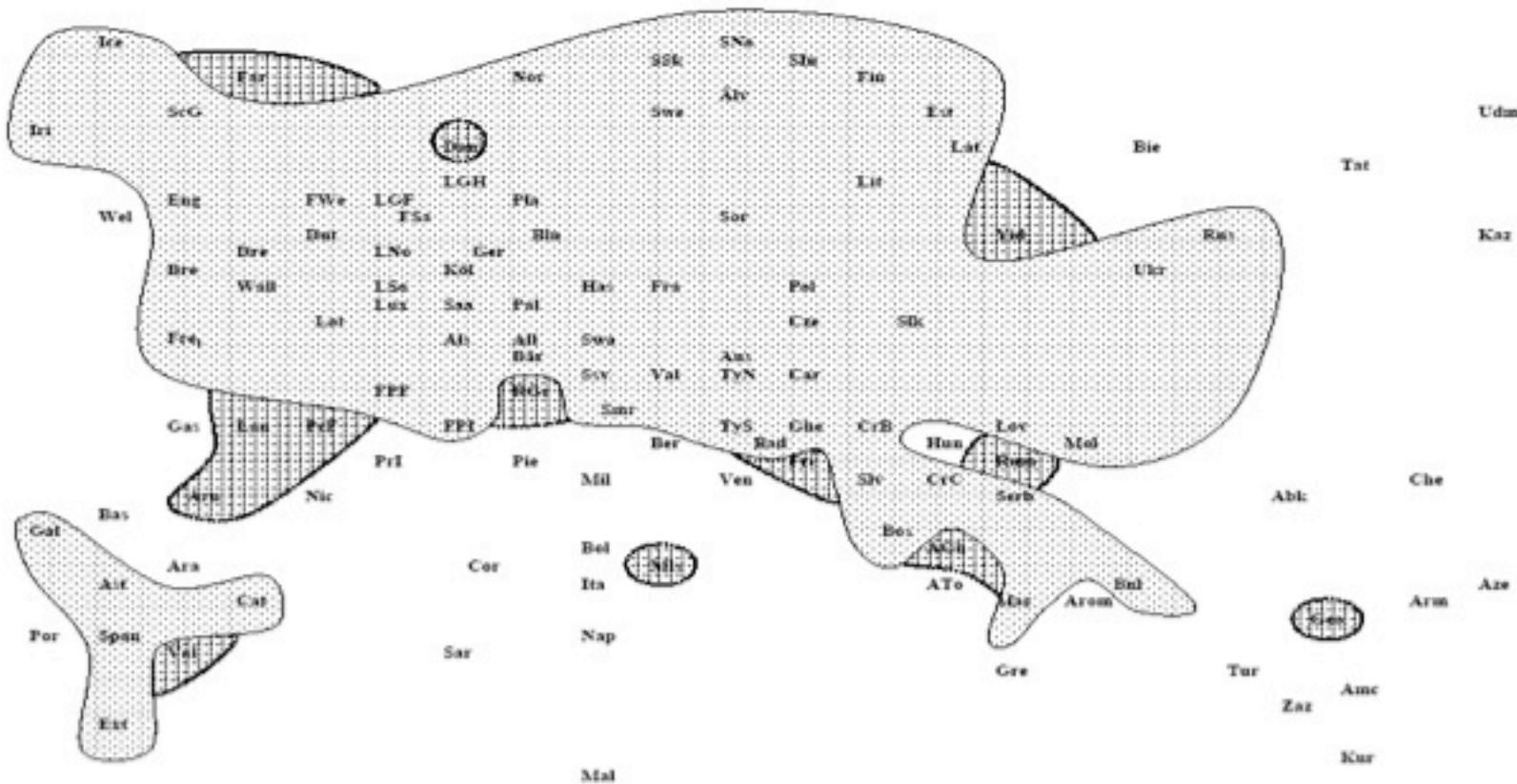
$n \geq 9$, $n \geq 6$ et $n \geq 3$

(C19) Type frequency $n \geq 9$, $n \geq 6$, $n \geq 3$



Types: Langues à fréquence minime et $n \leq 2$

(C20) Type frequency $n < 1$ and $n \leq 2$



Étude I: Relations tokens-type

- Les relations entre tokens et types dans les langues d'Europe donnent des résultats intéressants:
 - Ils confirment que les langues germaniques et d'autres langues à aversion pour le redoublement constituent un groupe à part.
 - En outre, ils indiquent une bipartition des langues à préférence pour le redoublement:
 - les langues non-indo-européennes s'accumulent au milieu du « classement »,
 - les autres langues – notamment les langues romanes – occupent des positions plus hautes.
 - Ca signifie que les dernières utilisent un nombre très limité de types avec une fréquence plus élevée quoique les langues non-indo-européennes se servent de plusieurs types avec une fréquence modérée.

(C21) Type-token ratios in *Le Petit Prince*

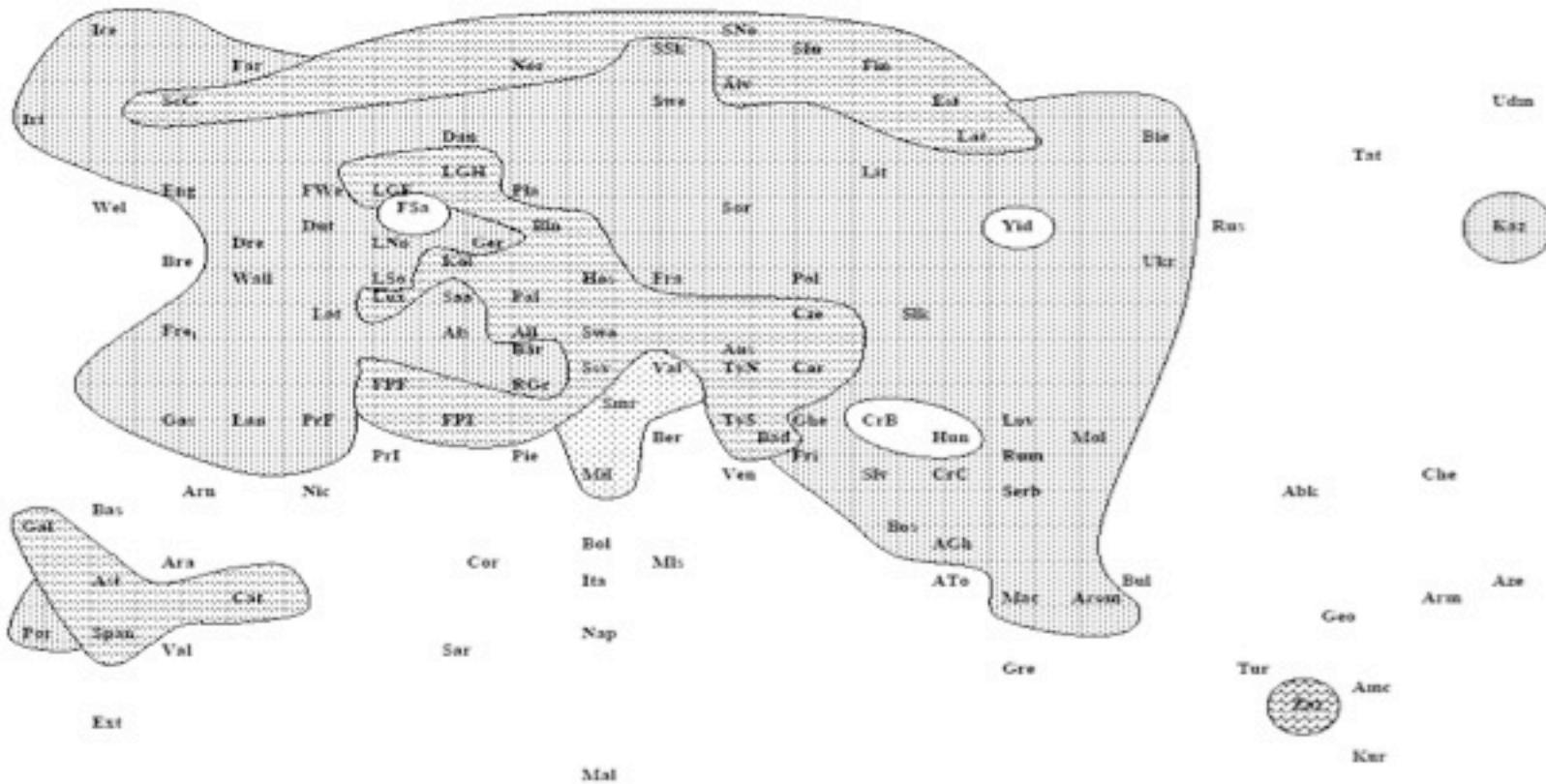
rank	language	tok/typ	typ/tok
1	<i>Niçard</i>	6.00	0.16
2	<i>Provençal (Italy)</i>	5.66	0.17
3	<i>Aranese</i>	4.50	0.22
4	Molisano	4.00	0.25
5	Maltese	3.28	0.30
6	Aramaic (East)	3.25	0.30
7	Georgian	3.00	0.33
8	<i>Corsican</i>	2.66	0.37
9	<i>Valenciano</i>	2.50	0.40
10	<i>Bolognese</i>	2.33	0.42
11	<i>Albanian (Tosk)</i>	2.16	0.46
12	<i>Napoletan</i>	2.05	0.48
13-17	<i>Croatian (Burgenland), Sardinian, Extremoñio, Ensian (Saterland), Russian</i>	2.00	0.50
18	<i>Venetian</i>	1.83	0.54
19	Udmurt	1.80	0.55
20	Greek	1.57	0.63
21	Basque	1.51	0.65
22-24	<i>Chechen, Italian, Yiddish</i>	1.50	0.66
25	Turkish	1.46	0.68
26-28	<i>Abkhaz, Aragonese, Welsh</i>	1.40	0.71
29	Hungarian	1.37	0.72
30-33	<i>Azeri, Bergamasco, Piemontese, Kurmancî</i>	1.33	0.75
34	Tatar	1.31	0.76
35	Armenian (East)	1.30	0.76
36	Zazaki	1.20	0.83
37-86	Albanian (Gheg), <i>Aromunian, Alsatian, Bärndüütsch, Bielarusian, Bosnian, Breton, Bulgarian, Croatian (Croatia), Danish, Dreents; Dutch; English, Faroese, Franconian; French, Frisian (West); Frisian, Gascon, German, Gherdëma; Icelandic; Irish; Kazakh, Languedocien, Milanese, Limburgian (Northern); Limburgian (Southern); Lithuanian; Lothringian; Lovari; Macedonian; Moldavian; Plautdietsch; Polish; Portuguese, Provençal (France), Romanian, Rumanian Grischun, Saami (Skolt); Saarlandash; Serbian; Slovak; Slovenian; Sorbian (Upper); Surmeir; Swedish; Ukrainian; Vallader, Wallon</i>	1.00	1.00

Cartographie III: Relations tokens-type

- Les deux cartes suivantes confirment encore une fois la distribution géolinguistique du phénomène:
 - Europe est subdivisé en deux aires majeures:
 - le nord, l'ouest et le centre – caractérisés par valeurs minimes et
 - le sud et l'est où les valeurs sont sensiblement plus hautes.

Rélations tokens-type minimes

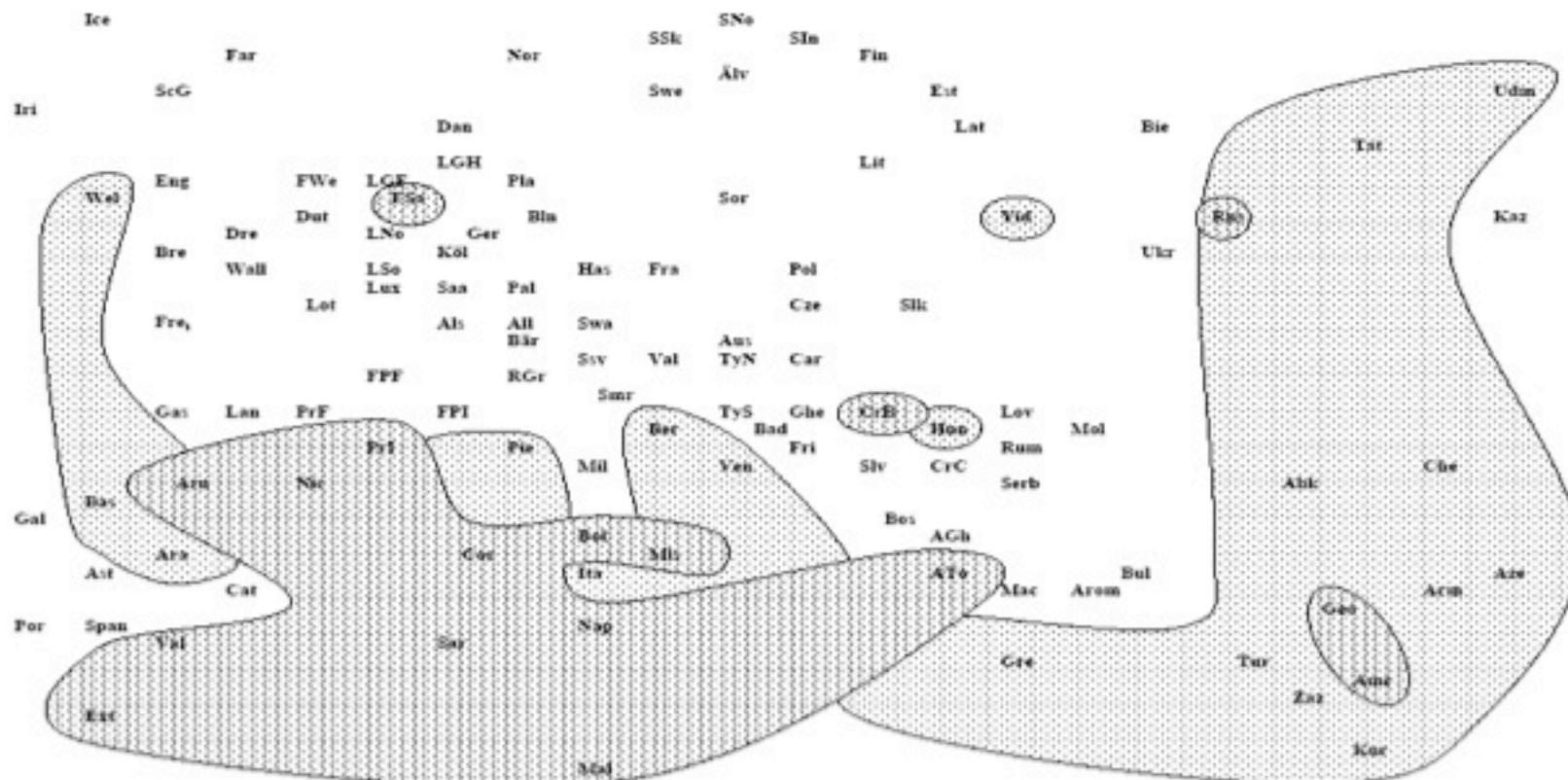
(C22) Token – type ratio: zero frequency and n = 1



Rélations tokens-type

$n \geq 2$ et $n \geq 1,18$

(C23) Token – type ratio: $n > 2$ and $n > 1.18$



Étude II: Tokens et types dans Harry Potter

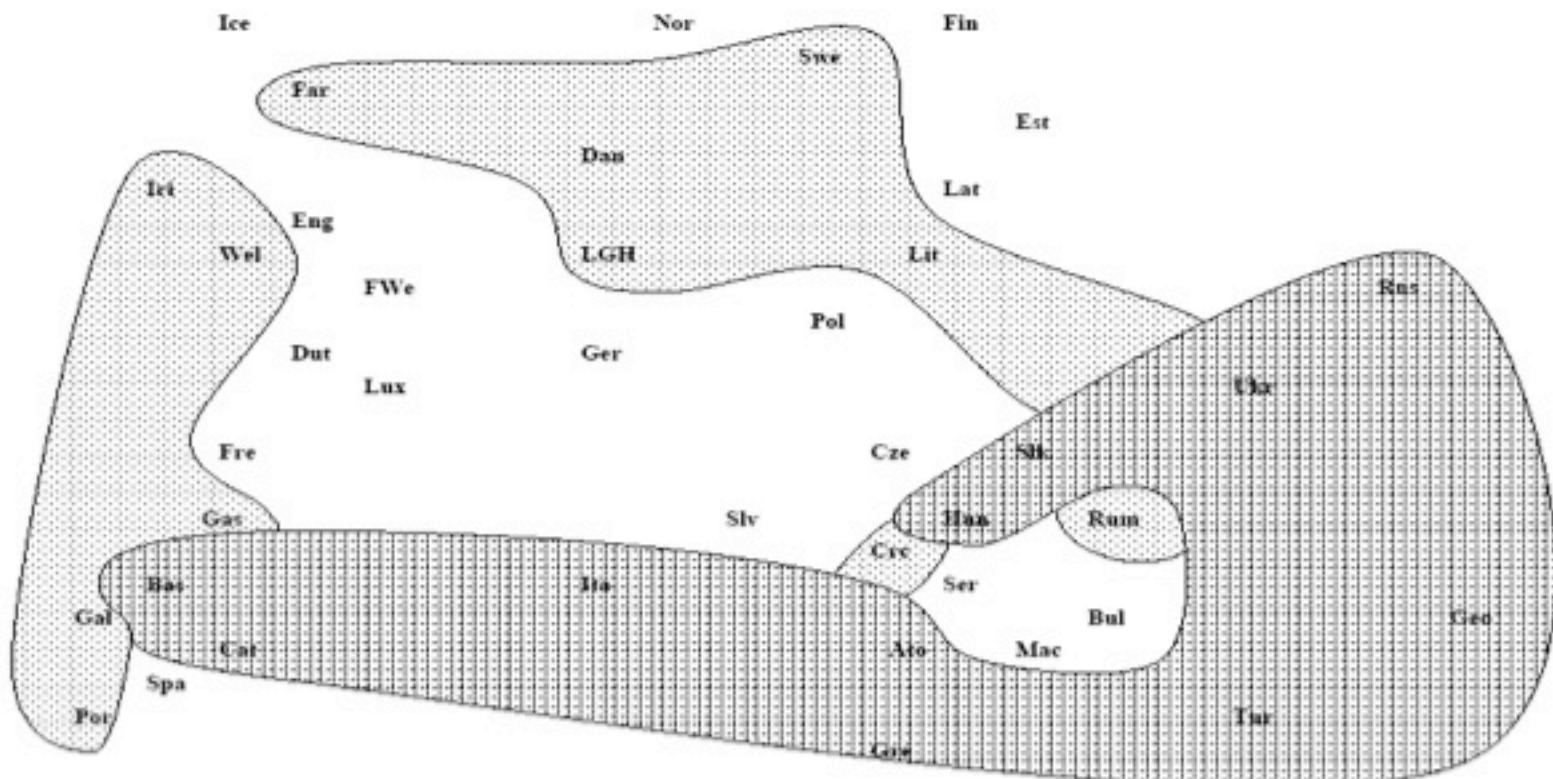
- L'analyse statistique de « Harry Potter, tome I » donne des résultats analogues:
 - **Tokens et types:** opposition entre langues non-indo-européennes (à préférence pour le redoublement) et les langues germaniques (à aversion pour le redoublement)
 - **Géolinguistique:** le centre d'Europe est encerclé par des langues qui se servent du redoublement relativement souvent.

(C25) Tokens of TR in *Harry Potter*

rank	language	tokens	deviation from estimates	
			sentences	words
1	Basque	316	+8	-256
2	Turkish	83	-50	-164
3	Georgian	58	+16	-18
4	Italian	29	-13	-29
5	Catalan	26	+26	+26
6	Russian	23	+9	-3
7	Hungarian	20	-57	-123
8	Slovak	19	+12	+6
9	Greek	16	-61	-127
10	Ukrainian	15	+8	+2
11	Albanian (Tosk)	13	-78	-156
12-13	Irish, Lithuanian	12	+5	-1
14	Welsh	12	-37	-79
15-16	Gallego, Low German (East Holstein)	10	+10	+10
17	Swedish	9	+2	-4
18-20	Danish, Faroese, Romanian	8	-6	-18
21	Gascon	8	-13	-31
22	Croatian (Croatia)	6	-1	-7
23	Portuguese	6	-15	-33
24-25	English, Serbian	5	-2	-8
26-27	Larvian, Norwegian (Bokmål)	4	+4	+4
28	Slovenian	4	-3	-9
29-31	Bulgarian, French, Frisian (West)	3	-4	-10
32-34	Czech, Finnish, Spanish	2	+2	+2
35	German	2	-5	-11
36	Luxembourgish	1	+1	+1
37-39	Dutch, Icelandic, Macedonian	1	-6	-12
40	Estonian	0	0	0
41	Polish	0	-7	-13
Total		770	-175	-1,154
Average		18.78	-4.26	-28.14

Tokens: Langues à fréquence $n \geq 13$ et $n \geq 6$

(C32) Token frequency $n \geq 13$ and $n \geq 6$

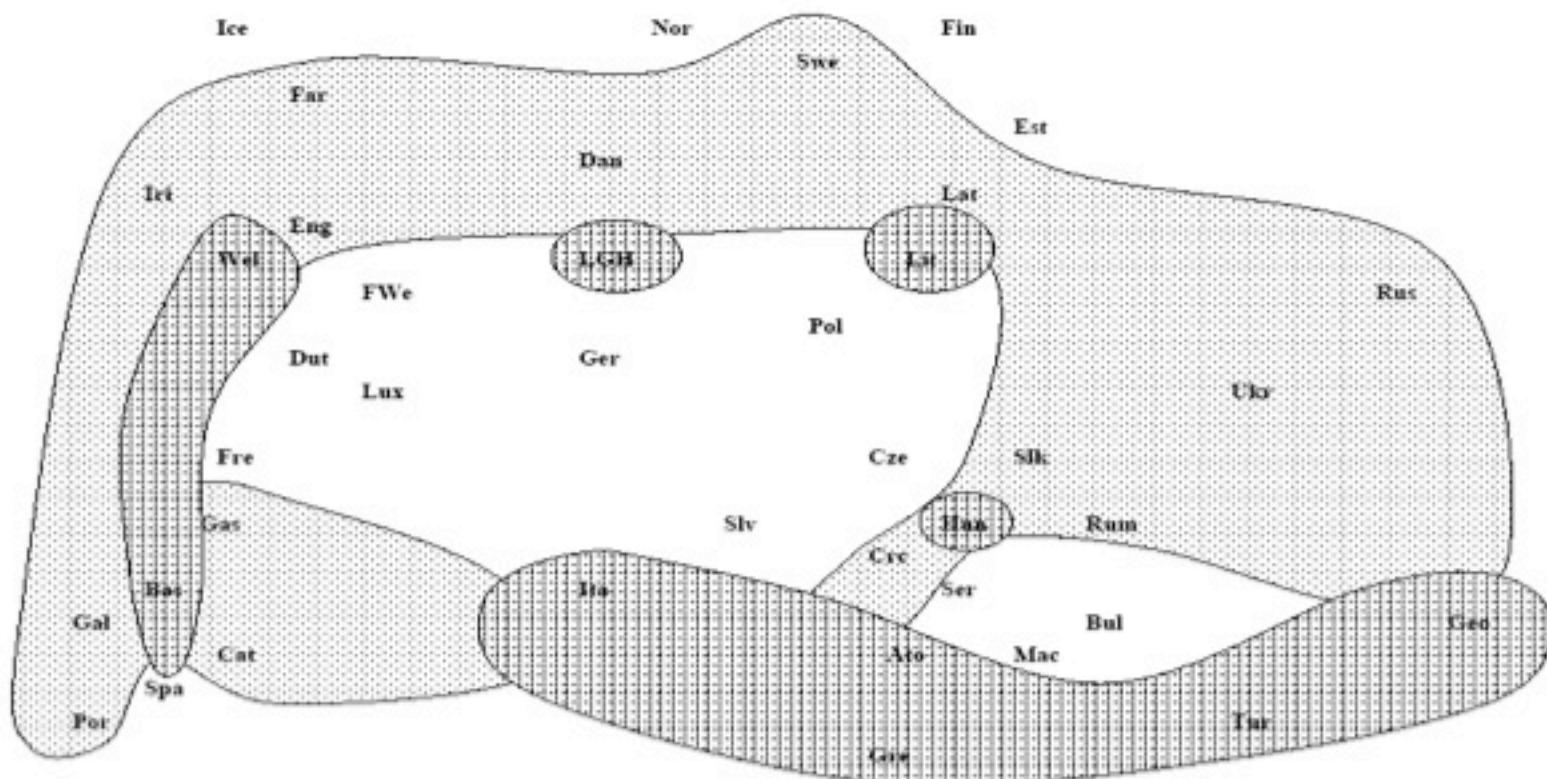


(C27) Types of TR in *Harry Potter*

rank	language	types	deviation from estimates	
			sentences	words
1	Basque	69	-134	-308
2	Turkish	51	-40	-118
3	Georgian	28	+14	+2
4	Italian	19	-9	-33
5	Welsh	11	-24	-56
6	Albanian (Tosk)	11	-31	-67
7	Greek	10	-39	-81
8	Lithuanian	9	+2	-4
9	Low German (East Holstein)	8	+8	+8
10	Hungarian	8	-48	-96
11-12	Slovak, Swedish	7	0	-6
13	Romanian	7	-7	-19
14	Galego	6	+6	+6
15-16	Irish, Russian	6	-1	-7
17	Faroese	6	-8	-20
18	Gazcon	6	-15	-33
19	Catalan	5	+5	+5
20-22	Croatian (Croatia), English, Ukrainian	5	-2	-8
23-24	Latvian, Norwegian (Bokmål)	4	+4	+4
25	Danish	4	-10	-22
26	Portuguese	4	-17	-35
27-29	Bulgarian, Frisian (West), Serbian	3	-4	-10
30-32	Czech, Finnish, Spanish	2	+2	+2
33-35	French, German, Slovenian	2	-5	-11
36	Luxembourgish	1	+1	+1
37-39	Dutch, Icelandic, Macedonian	1	-6	-12
40	Estonian	0	0	0
41	Polish	0	-7	-13
Total		336	-392	-1,016
Average		8.19	-9.56	-24.78

Types: Langues à fréquence $n \geq 8$ et $n \geq 4$

(C33) Type frequency $n \geq 8$ and $n \geq 4$



Étude II: Combinaison

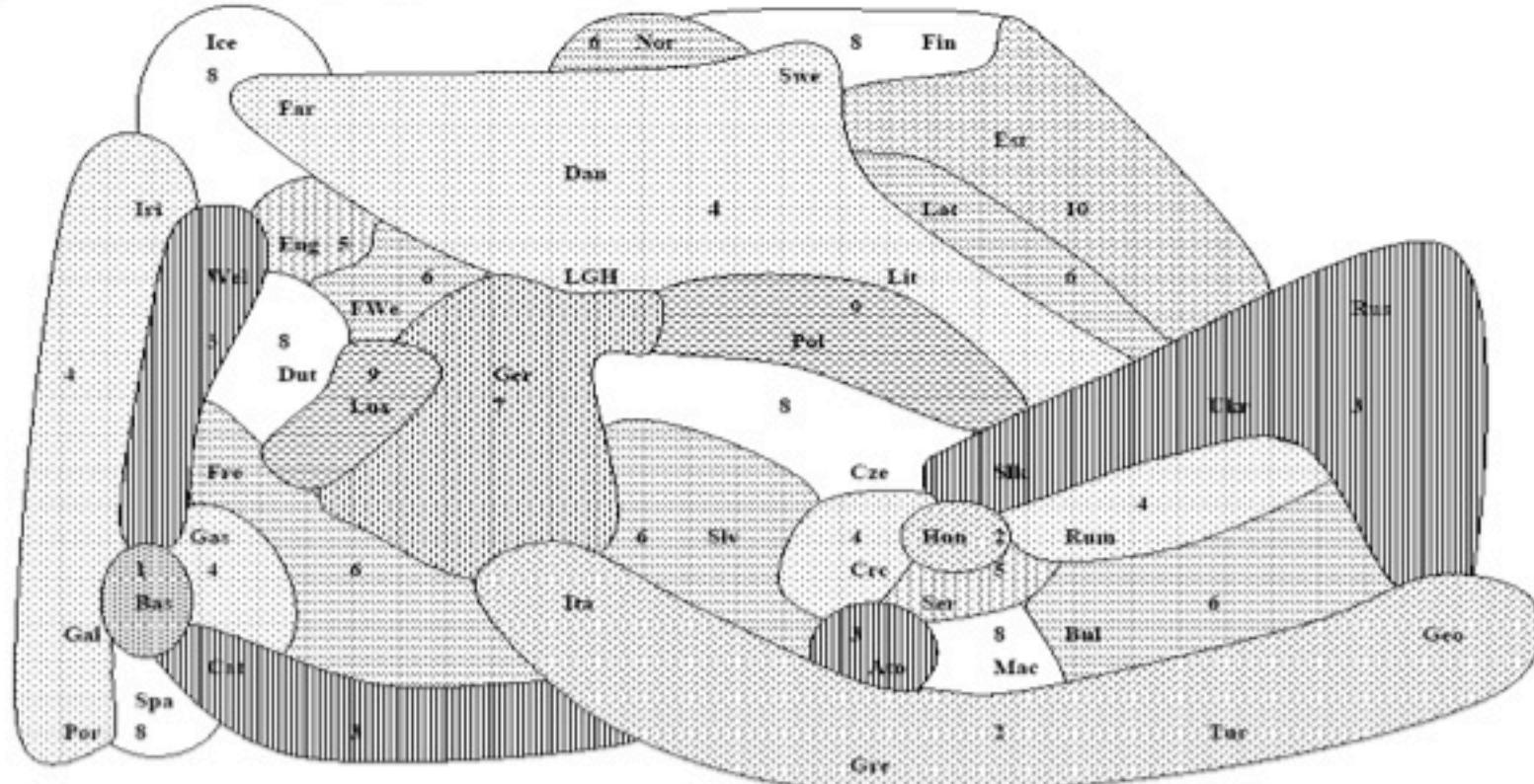
- On peut aussi combiner les résultats des deux études.
 - Comme ceux résultats sont presqu'identiques, leur combinaison ne peut que confirmer la tendance identifiée ci-dessus.

(C34) Combined token count

rank	language	tokens	share
1	Basque	360	27.10%
2	Turkish	102	7.68%
3	Georgian	64	4.81%
4	<i>Italian</i>	35	2.63%
5	Hungarian	31	2.33%
6	Greek	27	2.03%
7-8	Albanian (Tosk), <i>Catalan</i>	26	1.95%
9	Russian	25	1.88%
10	Slovak	20	1.50%
11	Welsh	19	1.43%
12	Ukrainian	16	1.20%
13-14	Irish, Lithuanian	13	0.97%
15	<i>Gascon</i>	11	0.82%
16-21	Danish, Faroese, <i>Gallego</i> , Low German (East Holstein), Rumanian, Swedish	10	0.75%
22	<i>Portuguese</i>	9	0.67%
23	Croatian (Croatia)	7	0.52%
24-25	English, Serbian	6	0.45%
26	Slovenian	5	0.37%
27-31	Bulgarian, French, Frisian (West), Latvian, Norwegian (Bokmål)	4	0.30%
32	<u>German</u>	3	0.22%
33-38	Czech, Dutch, Finnish, Icelandic, Macedonian, Spanish	2	0.15%
39-40	Luxemburgish, Polish	1	0.07%
41	Estonian	0	0.00%
Total		1,328	100.00%
average		32.39	2.43%

Tokens (somme des deux études): fréquence diminuant

(C35) Decreasing token frequency (combined results)



Étude II: Hors d'Europe

- Est-ce que les langues d'Europe se comportent comme les langues à redoublement parlées hors d'Europe?
 - Afin de répondre à cette question, j'ajoute deux sondages statistiques de langues extra-européennes:
 - les 16 langues étudiées à base de leurs traductions de « Le Petit Prince » montrent une diversité semblable à celle trouvée en Europe:
 - Il y a des langues qui utilisent le redoublement avec une fréquence énorme qui s'oppose à des valeurs statistiques bien plus basses
 - les 6 langues étudiées à base de leurs traductions de « Harry Potter, tome I » confirment ce résultat.
 - L'indonésien donne des résultats exceptionnels; plusieurs langues extra-européennes surpassent la majorité des langues européennes; il y a des langues extra-européennes qui ne tiennent pas avec les langues européennes à préférence pour le redoublement.
 - Donc, la pratique de Rubino (2005a) d'accepter les langues d'hors d'Europe presque toujours comme représentants du type de langues à redoublement tandis qu'il exclut la majorité des langues européennes de cette classe n'est pas du tout justifiée.

(C36) Token frequency and type frequency in *Le Petit Prince*
 (Control sample)

rank	language	tokens	types	ratio
1	Indonesian	395	132	2.99
2	Hakka	112	71	1.57
3	Tagalog	55	33	1.66
4	Morisien	28	19	1.47
5	Vietnamese	27	20	1.35
6	Kirgiz	26	18	1.44
7	Quechua (Cuzco)	24	20	1.20
8	Bambara	20	10	2.00
9	Hebrew	19	10	1.90
10	Uzbek	18	14	1.28
11	Papiamentu	15	7	2.14
12	Zulu	6	2	3.00
13	Guarani	5	4	1.25
14	Kabyle	5	3	1.66
15	Afrikaans	2	2	1.00
16	Toba	0	0	0.00
Total		757	365	2.07
average		47.31	22.81	2.07

(C37) Token frequency and type frequency in
Harry Potter (Control sample)

rank	language	tokens	types	ratio
1	Indonesian	1,450	394	3.68
2	Vietnamese	104	56	1.85
3	Mongolian	62	35	1.77
4	Afrikaans	30	19	1.57
5	Hebrew	18	8	2.25
6	Greenlandic	4	2	2.00
Total		1,668	514	3.24

Étude II: Le long d'Oural

- L'image de l'Europe comme territoire sans redoublement change considérablement si – entre autres – les langues de l'Eurasie sont incluses dans le cadre de langues.
 - Dans la table suivante, on voit que langues comme l'oudmourte, le bachkir, le tchouvache, le kalmyk donnent des résultats relativement élevés pour les tokens et les types du redoublement (tous identifiés dans des textes originaux comparables à « Harry Potter, tome I »).

(C38) Statistics of further TR-languages in the European East

language	tokens	types	ratio tokens/type	ratio tokens/sentence	ratio tokens/word
Avar	289	119	2.42	0.051	0.003
Chuvash	245	146	1.67	0.041	0.003
Ossetic	137	60	2.28	0.028	0.002
Udmurt	127	81	1.56	0.046	0.003
Bashkir	93	48	1.93	0.016	0.001
Kalmyk	79	36	2.19	0.017	0.001

Études II: Relations types-frase

- La table suivante reflète la relation entre types et frase dans mes corpus divers.
 - La table spécifie cette relations pour 151 langues (y incluses celles d'hors d'Europe).
 - Les premières deux positions sont occupées par des langues extra-européennes.
 - Dans le reste du « classement », langues européennes et extra-européennes alternent (plus ou moins).
 - Les langues additionnelles – comme l'arabe de Chypre – se placent parmi celles à préférence pour le redoublement.

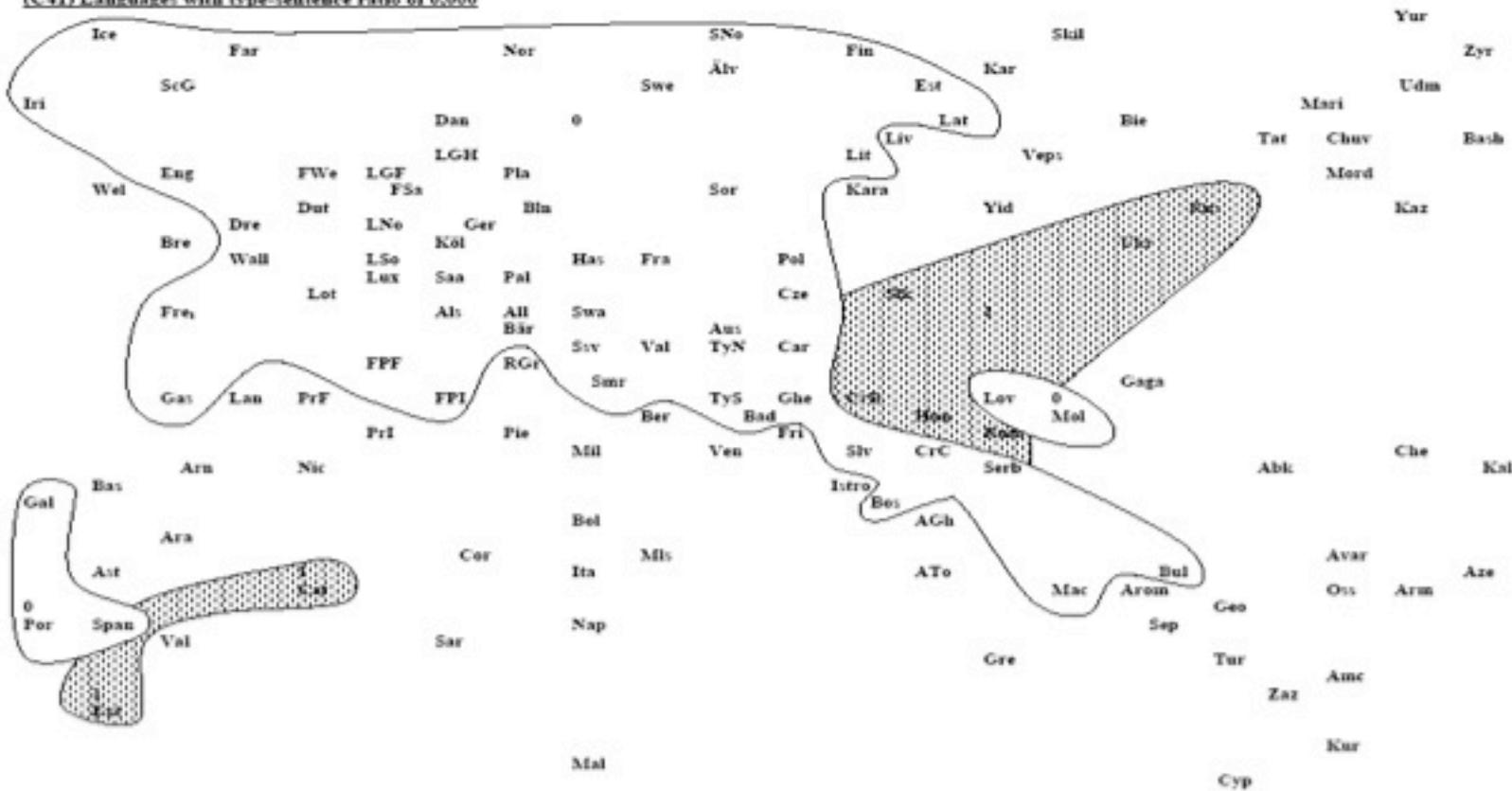
rank	CC40: Type-entourage ratio hierarchy - language	10^-38	10^-36
1	Catalan	0.067	0.043
2	Indonesian	0.139	0.035
3	Chuvash	0.041	0.024
4	Mordvin	0.032	0.023
5	Aver	0.051	0.021
6	Jasslog	0.033	0.020
7	Udmurt	0.032	0.019
8	Cypriot Arabic	0.041	0.016
9	Azer	0.021	0.016
10	Monisien	0.016	0.015
11	Veps	0.021	0.014
12	Ossetic	0.028	0.012
13-14	Mari, Quechua (Cuzco)	0.014	0.012
15	Tatar	0.015	0.011
16	Napoletan	0.031	0.010
17	Kurziz	0.015	0.010
18	Chechen	0.014	0.009
19	Abkhaz	0.012	0.009
20	Saami (Kildin)	0.010	0.009
21	Bashkir	0.016	0.008
22	Uzbek	0.010	0.008
23	Kalmyk	0.017	0.007
24	Albanian (Obeg)	0.013	0.007
25	Livonian	0.012	0.007
26	Basque	0.037	0.006
27	Basotho	0.012	0.006
28	Armenian (East)	0.007	0.006
29	Septardic	0.016	0.005
30	Corsican	0.014	0.005
31	Sardinian	0.010	0.005
32	Vietnamese	0.009	0.005
33	Persian	0.009	0.004
34	Turkish	0.007	0.004
35	Karelian	0.008	0.004
36	Venetian	0.006	0.003
37	Mongolian	0.005	0.003
38	Aragonese	0.004	0.003
39-41	Milanese, Belarusian, Zazaki	0.003	0.003
42	Armenian (East)	0.007	0.002
43	Georgian	0.004	0.002
44-47	Breton, Guaraní, Kurmandji, Zulu	0.003	0.002
48-50	Arenjumian, Istro-Romanian, Yorùbá	0.002	0.002
51-52	Nogard, Provençal (Italy)	0.010	0.001
53	Maltese	0.007	0.001
54	Armenia	0.005	0.001
55-56	Molisano, Bolognese	0.004	0.001
57-58	Kabyle, Valenciano	0.003	0.001
59-64	Afrikaans, Bergamasco, Greek, Italian, Piemontese, Welsh	0.002	0.001
65-71	Albanian (Tosk), Friulan, Kazakh, Languedocien, Provençal (France), Rumanian, Grischun, Yiddish	0.001	0.001
72	Hungarian	0.005	0.000
73	Hebrew	0.003	0.000
74-81	Bosnian, Catalan, Croatian (Burgenland), Extremeno, Rumanian, Russian, Slovak, Ukrainian	0.001	0.000
82-151: Allemannic (Baden), Alpedalsk, Alemann, Austrian (Upper), Badische, Baudeutsch, Berliner, Bulgarian, Carinthian, Croatian (Croatia), Czech, Danish, Dreants, Dutch, English, Estonian, Faroese, Finnish, Francian, Franco-Provençal (France), France-Provençal (Val d'Aoste), French, Frisian (Friesland), Frisian (West), Gallego, Gascon, German, Gherdëina, Graanländer, Hassika, Icelandic, Irish, Kóisc, Larzac, Limburgian (Northern), Limburgian (Southern), Lithuania, Lotharingian, Lower, Low German (East Frisia), Low German (East Holstein), Luxembourgish, Macedonian, Moldavian, Norwegian (Bokmål), Palauan, Piedmonteñ, Polish, Portuguese, Saami (Inari), Saami (Northern), Saami (Skolt), Saarlandish, Saxon-Gothic, Sardinian, Slovensian, Sorbian (Upper), Spanish, Surmuri, Swiss-German, Swabian, Swedish, Taib, Tyrolese (Northern), Tyrolese (Southern), Vallader, Wallon, Zwyvyan 0.000 - 0.000			

Étude II: Géolinguistique des relations types-frase

- Les deux cartes complétées suivantes terminent ma présentation.
 - Elles démontrent que la relation types-frase se distribue selon des critères géolinguistiques qui correspondent à ceux mentionnées ci-dessus:
 - les relations basses se trouvent dans le nord, l'ouest et le centre,
 - les relations à valeurs hautes occupent les régions plus au Sud et à l'Est.

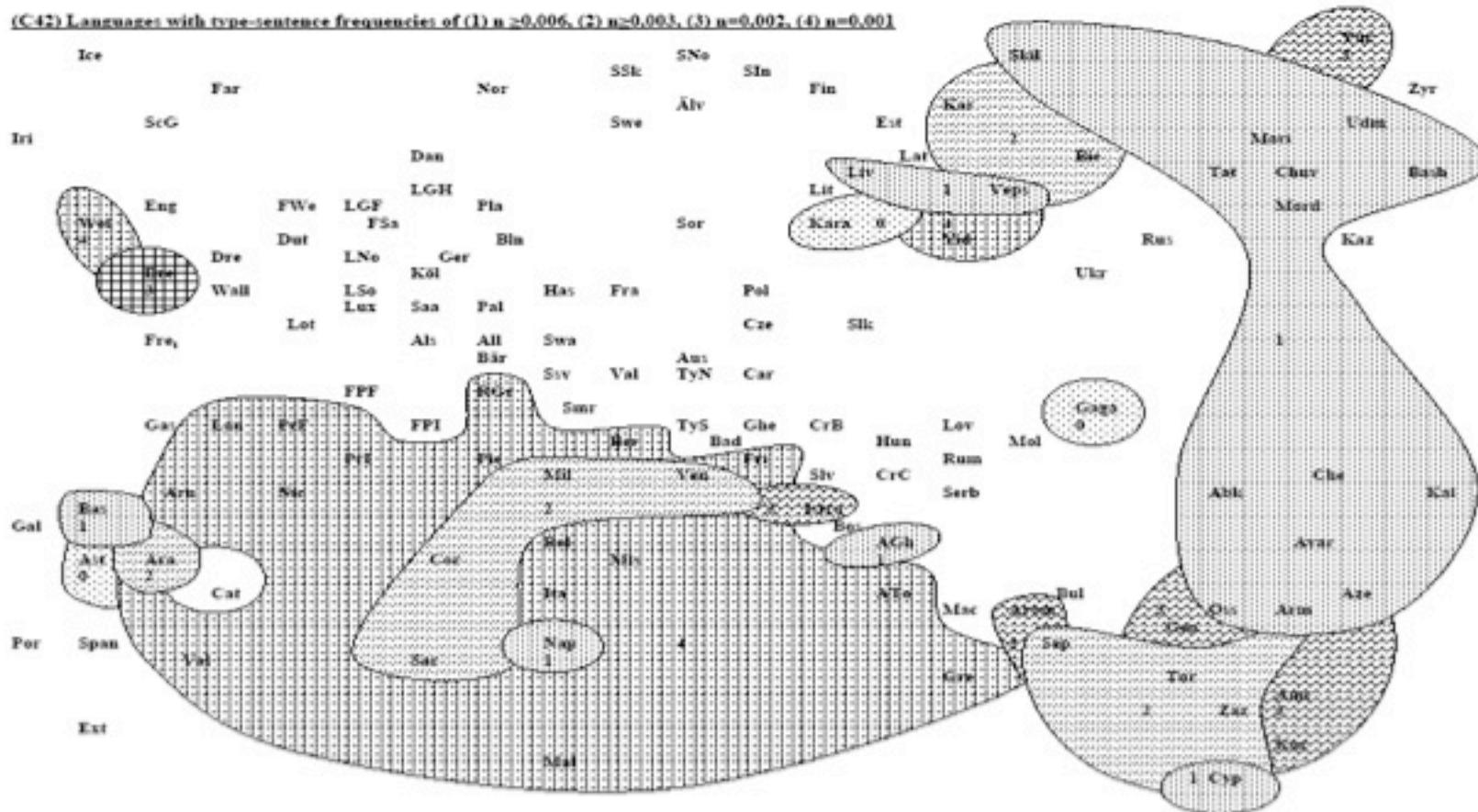
Langues à relation types-phrase zéro

(C4) Languages with type:sentence ratio of 0.000



Langues à relation types-phrases plus élevée

(C42) Languages with type:sentence frequencies of (1) $n \geq 0.006$, (2) $n \geq 0.003$, (3) $n = 0.002$, (4) $n = 0.001$



La Fin

- J'ai démontré que
 - les langues d'Europe ne peuvent pas toutes être traitées de la même manière:
 - il y a des langues européennes qui se servent fréquemment du redoublement et
 - il y a d'autres qui évitent le redoublement,
 - les deux groupes de langues se distribuent selon des critères géolinguistiques très claires:
 - le nord, l'ouest et le centre forment une aire compacte où le redoublement est un phénomène (presque) inconnu,
 - le sud et l'est sont dominés par les langues à préférence pour le redoublement.
 - Alors, Haspelmath (2001) a le bon droit de son côté.

Encore la fin

- La diversité des résultats statistiques en Europe n'est pas trop différente de celle attestée hors d'Europe.
- C'est pourquoi il n'est pas justifié d'exclure les langues européennes de la classe des langues à redoublement total.
- Afin de vérifier si cette révision de la typologie proposée par Rubino (2005a) est correcte, il faut développer une mesure statistique universelle de la productivité.
- Au moins les régions orientales et méridionales d'Europe correspondent plus à l'image du redoublement qu'on trouve hors d'Europe.