

Grondelaers, Stefan, Dirk Geeraerts, Dirk Speelman & José Tummerts. 2001. "Lexical standardisation in internet conversations. Comparing Belgium and The Netherlands". In Josep M. Fontana, Louise McNally, M. Teresa Turell & Enric Vallduví (red.), *Proceedings of the First International Conference on Language Variation in Europe* 90-100. Barcelona: Universitat Pompeu Fabra, Institut Universitari de Lingüística Aplicada, Unitat de Investigació de Variació Lingüística.

LEXICAL STANDARDISATION IN INTERNET CONVERSATIONS: COMPARING BELGIUM AND THE NETHERLANDS

Stefan Grondelaers
Dirk Geeraerts
Dirk Speelman
José Tummerts

University of Leuven

1. Introduction

This paper reports on ongoing research into diachronic and synchronic aspects of the linguistic relationship between Belgian and Netherlandic Dutch. The investigation basically addresses the question whether Belgian Dutch – as spoken in Flanders, the Belgian region south of The Netherlands – converges with the variant of Dutch spoken in The Netherlands, or whether it represents an autonomous language variety. This descriptive question is closely linked to a controversial language political debate: should Dutch speaking Belgium adopt Netherlandic Dutch as its linguistic norm (the "integrationist" view), or is there room for an official Belgian variant of the Dutch standard (the "particularistic" view)? We will concern ourselves exclusively with the descriptive issue in this paper.

The reader who is not familiar with the history of Belgian Dutch should recall that in the Renaissance, linguistic standardisation was well on its way in most European countries. In Flanders, however, ongoing standardisation was blocked as a result of Flanders' political separation from The Netherlands in the Eighty Years' War. Instead of developing its own linguistic standard, Flanders adopted French for supraregional purposes, especially after the birth of the Belgian state in 1830. Although Dutch was reinstated as Flanders' official language, and although it resumed its standardisation (with a major leap after World War II), linguists agree that the process is not yet finished.

With respect to the status and the development of Belgian Dutch, two uncontroversial hypotheses can be found in the linguistic literature. First, there is an expectation of *diachronic convergence* between Belgian and Netherlandic Dutch: the standardisation of Belgian Dutch is characterised by a progressive normative orientation towards Netherlandic Dutch. In addition, the unfinished character of the standardisation of Belgian Dutch is believed to manifest itself in a larger *synchronic distance* between local and national language in Belgium than in The Netherlands; even to the untrained observer, it is obvious that the differences between regional and supraregional register are much larger in Belgium than in The Netherlands. We will henceforward refer to both hypotheses as the *diachronic* and the *synchronic* hypothesis respectively.

Since, unfortunately, neither the diachronic nor the synchronic hypothesis is based on more than anecdotal observation or questionnaire-elicited language data (which need not correspond to actual language use, cf. Labov 1972, Geeraerts, Grondelaers and Speelman 1999: 24 and Grondelaers 2000: 59-63 and 256), a research team lead by Dirk Geeraerts launched a large-scale usage-based investigation into the lexical standardisation of Belgian Dutch, using a 40.000 item database of synonymous designations of 30 noun-concepts from two lexical fields, viz. clothing and football. In the database, a *geographic* as well as a *diachronic* dimension was implemented, by contrasting Belgian and Netherlandic Dutch from 1950, 1970, and 1990. In addition, we integrated a *stylistic-stratificational* dimension in the database, on which supraregional language use from national magazines was compared to regional language data recorded from the price tags in the windows of clothing boutiques in two Belgian and two Netherlandic towns. The principal results of this investigation are briefly summarised in section 2 of this paper (a detailed overview can be found in Geeraerts, Grondelaers & Speelman 1999). Section 3 reports on a new replication study carried out to corroborate the synchronic findings in Geeraerts, Grondelaers & Speelman (1999). The new study specifically explores the possibilities of Internet language as an alternative for the regional shopwindow material in the first study, concentrating not only on IRC – Internet Relay Chat – but also on the language of Usenet conversations. In section 4, we will elaborate on the nature and origin of the discrepancies between the original data and the replication findings, and we will briefly discuss subsequent directions in the research.

2. A usage-based approach to the standardisation of Belgian Dutch

2.1. Materials

The empirical foundation of the initial research project consisted of 40.000 observations of language use. We collected the different names (and their frequencies) which are used to denote 30 concepts, 15 from the field of clothing terminology, and 15 from the field of soccer terminology. The resulting database allows us, for instance, to calculate the proportion in Belgian and Netherlandic sources of the term *buitenspel* ‘offside’ and the loanword *offside* for the concept OFFSIDE; in the case of the concept JURK ‘dress’, we can determine whether the lexical choices involve a preference for either *jurk*, *japon*, or *kleed*. The core of the observed material consisted of magazine- and newspaper material recorded in 1990; this core was extended in two ways.

In the first place, similar material was collected for 1950 and 1970, which enabled us to carry out a ‘real time’-investigation of lexical convergence or divergence processes. In addition, the stratification of language use was taken into account. Between standard and dialect, there are a number of ‘strata’ on which register differences may cooccur with an increasing geographical specialisation. For an investigation of the relationship between Belgian and Netherlandic Dutch, these strata – viz. the regionally coloured informal variants of the standard language – are extremely relevant: it can be expected that the linguistic differences between Belgium and The Netherlands will increase on this regiolectic level.

In the initial investigation, the intermediate level between dialect and written standard language was represented by the clothing terms¹ we collected from labels and price tags in shopwindows in two Belgian (Leuven and Kortrijk) and two Netherlandic towns (Leiden and Maastricht). The intended audience of this form of communication is more restricted than the national or binational audience which is the target of the magazines from which the core material was selected. The fact that we are dealing with written language in a semi-formal situation, on the other hand, ensures that we steer clear of the purely dialectical pole of the stratificational continuum.

2.2. Methods

How does one quantify lexical convergence or divergence between two language varieties? In Geeraerts, Grondelaers & Speelman (1999), a measure of lexical overlap was developed, based on the notions *onomasiological profile* and *uniformity*.

The *onomasiological profile* of a concept in a particular source is the set of synonymous names for that concept in that particular source, differentiated by relative frequency. Table 1 contains the onomasiological profiles for OVERHEMD ‘shirt’ in the Belgian and the Netherlandic 1990-database:

| | B90 | N90 |
|----------|------|------|
| hemd | 31 % | 17 % |
| overhemd | 69 % | 46 % |
| shirt | 0 % | 37 % |

Table 1: onomasiological profiles for SHIRT in the Belgian and Netherlandic 1990-database

Uniformity is a measure for the correspondence between two onomasiological profiles. Our computation of uniformity has its starting-point in the idea that a common language norm triggers uniform linguistic behaviour. In its most extreme form, lexical uniformity in the naming of a concept obtains when two language varieties have an identical name for that concept, or several names with identical frequencies in the two varieties. Much more frequent than these examples of “ideal” uniformity, however, are such partial correspondences as illustrated in table 1. Let us, for the sake of illustration, assume that the relative frequencies in table 1 represent 100 actual naming instances in each of both profiles, rather than percentages. The partial overlap between the profiles in table 1 is quantified by counting the naming instances for which there is a counterpart in the other profile. In the ideal scenario outlined above, each of the 100 naming events in each of both profiles has its counterpart in the other profile, yielding a maximal uniformity of 100 %. In table 1, however, 14 instances of *hemd* in B90 have no counterpart in N90, 23 Belgian *overhemden* have no Netherlandic counterpart, and there are no Belgian counterparts for the 37 Netherlandic *shirts*. On the grand total of 200 naming events in the two profiles, only $200 - (14 + 23 + 37) = 126$ instances have counterparts in the other profile, which yields a uniformity of $126/2 = 63$ %. For quantitative convenience's sake, it should be noticed that this percentage equates the sum of the smallest relative frequency for each alternative term, i.e. $17 + 46 + 0 = 63$ %.

If more than one concept is investigated, uniformity index U is defined as the average of the uniformity indexes of the separate concepts, whereas uniformity index U' is defined as a *weighted average*, in which the relative frequency of each concept in the investigated samples is taken into account. In this paper, we will focus exclusively on the weighted uniformity U' , in which high frequency concepts have a more outspoken impact on the overall uniformity.

2.3. Results

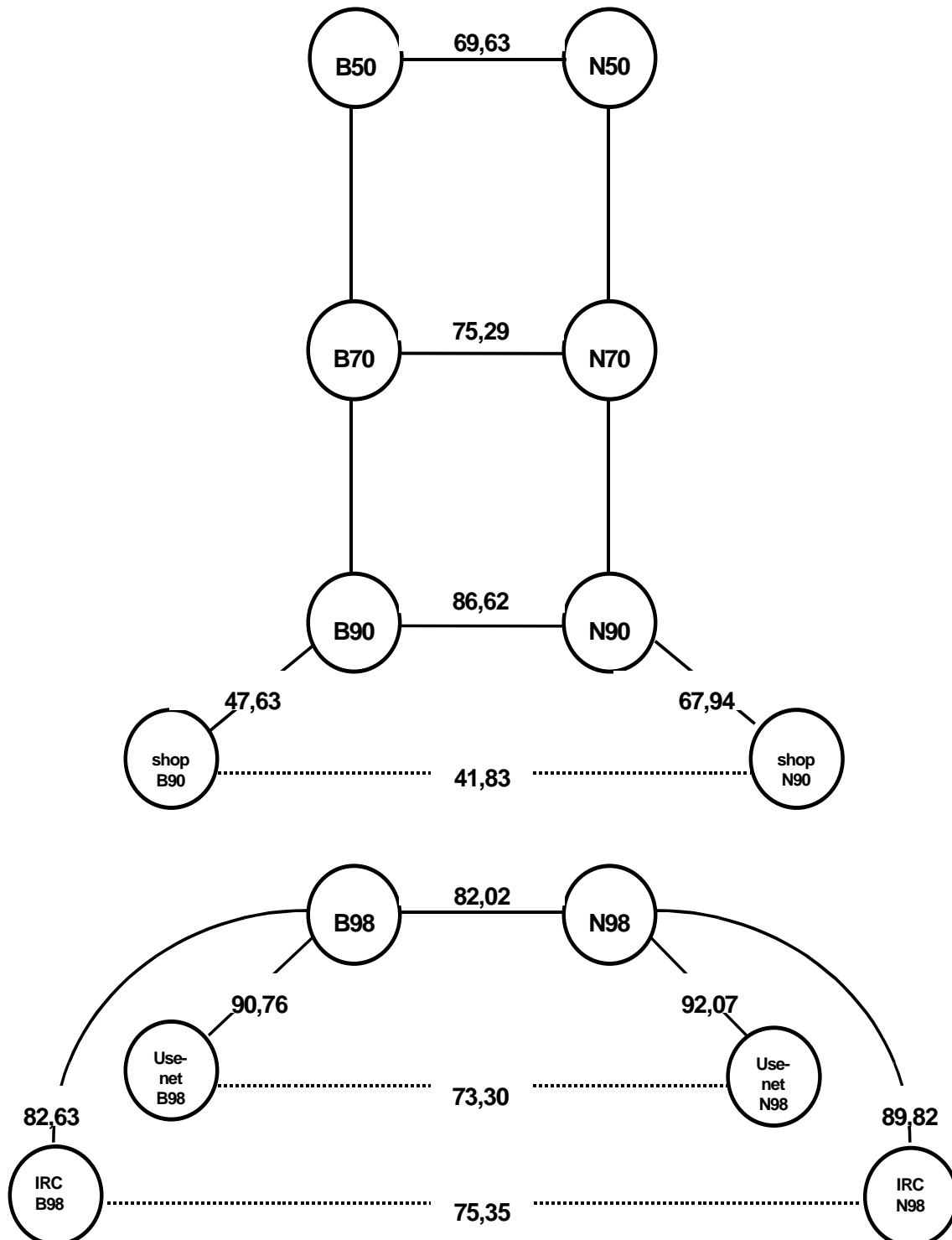


Figure 1: Uniformity between Belgian and Netherlandic 1950-, 1970-, 1990-, 1998-, shopwindow-, IRC-, and Usenet-samples

Let us, before discussing the results, operationalise the diachronic and the synchronic hypothesis in terms of uniformity. Diachronically, convergence and divergence can be quantified as increasing or decreasing uniformity. Synchronically, the larger distance between national and local language we expect in Belgian Dutch, will manifest itself in a smaller uniformity between magazine and shopwindow material in Belgian Dutch than in Netherlandic Dutch.

The top part of figure 1 contains the relevant uniformity values. The left axis in the figure represents the Belgian (B) material, whereas the right axis corresponds to the Netherlandic (N) data (“50”, “70”, “90” and “shop” stand for the 1950-, 1970-, 1990-, and shopwindow-material respectively). The value on the lines between the circles represents the weighted uniformity between the samples in the circles.

The data in figure 1 unambiguously confirm the diachronic as well as the synchronic hypothesis. Diachronically, the increase in uniformity between Belgian and Netherlandic Dutch suggests an evident lexical convergence between both varieties:

$$\begin{array}{rcl} U'(B50,N50) < & U'(B70,N70) < & U'(B90,N90) \\ 69.63 & < & 75.29 & < & 86.62 \end{array}$$

Synchronically, the delayed standardisation of Belgian Dutch manifests itself in a distinctly lower uniformity between the Belgian magazine and shopwindow data than between the Netherlandic magazine and shopwindow material:

$$\begin{array}{rcl} U'(B90,shop) < & U'(N90,shop) \\ 47.63 & < & 67.94 \end{array}$$

3. Replicating the synchronic findings with Internet language

3.1. Materials

Although the data in figure 1 constitute a solid confirmation of the diachronic and synchronic hypothesis, our case is not yet watertight. In addition to Van de Velde’s (1996) findings, which demonstrate that there is a marked *divergence* between Belgian and Netherlandic Dutch as far as pronunciation is concerned, there are shortcomings in our investigation which concern the dependent as well as the independent variables of the research. First, the empirical basis on which our conclusions are founded is too small: more linguistic parameters are needed than two lexical fields before we can safely state anything about the internal structure of Belgian Dutch (we will briefly address this issue in the final paragraph). Second, our implementation of the stylistic parameter does not do justice to the stratificational complexity of Belgian Dutch. Up to this point,

we have integrated only one intermediate language stratum in the research – represented by the shopwindow material -, but in the literature, two (Geerts 1975), three (Baetens Beardsmore & Van de Craen 1979), and even five (Willemyns 1987) distinct varieties of Belgian Dutch have been distinguished.

The questions we would therefore like to address in this section are: (i) can we replicate the synchronic findings of the first investigation with other types of informal language data than clothing names on shopwindow labels?, and (ii) do the new data - in combination with the original findings - reveal more about the internal structure of Belgian Dutch and, more particularly, about its stratification?

In order to answer these questions, a 45.000.000 word corpus of written Belgian and Netherlandic Dutch was compiled, which contains two types of informal Internet language as an alternative for the shopwindow data in the first investigation. *IRC 'Internet Relay Chat'*, the language of online chat channels, can be regarded as an example of “spoken language in written form”: although produced in a written medium, it shares with spoken language a dialogical immediacy that ordinary written language usually lacks. As a consequence, what is “written” in IRC tends to exhibit characteristics of (very) informal speech, with lexical and grammatical colloquialisms that are normally absent in ordinary written text (for an overview of the linguistic and sociological characteristics of Internet Relay Chat, cf. a.o. Ko 1996, Hentschel 1998, Bays 1998, Paolillo 1999, and especially Herring 1996). Example (1) contains an illustration of English IRC comparable to the Belgian and Netherlandic IRC-extracts in the corpus:

1. (ac22): ok, im 18, all of you r in your mid to late 20's, im going to college, while you r in a stupid chatroom, now lets see who is lame, not me, i would have to say you guys have to get a life.
(Morisato): cause they;d enjoy this
(@pai): they'd just spew on the keyboard
(Morisato): true :)
(@Cash): 20= 7300
(@pai): ac22- it looks like you're in a chatroom too, and i'm going to college too. boy that didn't work very well.
(ac22): ok, im a freshman, what r u
(Morisato): I dun go to college I work first to save for it
(@Awenyedd): hmm
(Morisato): costs money for college and all that
(@Awenyedd): hmmm
(@Awenyedd): well..let's see I'm a health care aid
(`JeSSiCa): u all r a bunch of fuckin retards and if any of u bitches thing about talkin to MY MAN i will kill u.. and talk shit to my boyfriend one more time i will beat the living hell out of each and every fuckin one of u.. u all are a bunch of stupid cock sucking mother fuckers and that need to get their ass kicked and that need to get some fuckin lives u stupid retarded bitches!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!

(posted on <http://www.egotrip.dk/logs/ac22.shtml>)

An interesting feature of Dutch IRC is the existence of regional antagonism between Belgian and Netherlandic users: Belgian channels typically remove Netherlandic users and vice versa (usually by means of a procedure which makes it impossible for persons with an *@*.be address to log in on a Netherlandic channel, and vice versa). For the linguist, this regional tension is beneficial, since it leads to relatively homogeneous corpora of spontaneously produced, maximally informal Dutch.

The use of IRC in linguistics is not, however, trouble-free. First, the appalling spelling makes digital extraction of the investigated phenomenon rather tricky. More fundamental disadvantages of IRC are the general absence of long sentences (and, hence, complex linguistic phenomena), and a frequent interruption of the discursive continuity which is the result of the fact that different conversations take place simultaneously. To compensate for these shortcomings, we have included a relatively larger portion of IRC in the new corpus.²

Usenet consists of offline³ discussion groups to which Internet users contribute by means of e-mail messages which are added to the ongoing debate. Example (2) contains an extract from an English Usenet message similar to the Belgian and Netherlandic Usenet messages in the corpus:

2. From: "John S. Dyson"
Newsgroups: comp.unix.bsd.openbsd.misc
Subject: Re: BSD licensing permits embrace, extend, and extinguish?
Lines: 39
Date: Fri, 09 Jun 2000 10:52:03 -0500

Luke Seubert wrote:

>
> "John S. Dyson" wrote:
>>
>> By degrading the argument into 'you can choose any license that you wish'
>> begs the claim that GPL is a license that controls the disposition of
>> derivative works. Those derivative works are the product of developers.
>> By limiting the freedom of developers, the users of the work product:
>> distributors and end-users get the advantage.
>>
> Are you sure? The benefit most sought by the end user is the
> availability of good software. What if some piece of abandoned GPL'd
> code in need of some work scares off a potential commercial developer,
> who might otherwise develop the code if it had a BSD type license?
>

That is indeed a case AGAINST the GPL... It makes sense in this case, if the changes to the GPLed works are trivial and the market isn't significant, then it is probably safe to work on the GPLed code (and charge appropriately for the opportunity cost of the GPLed works.) In fact, by using GPL code for works to be modified, then the cost to the original customer in this case (the one who pays and not the subsequent 'freeloaders') is likely higher.

(...)

John | Never try to teach a pig to sing,
dyson@iquest.net | it makes one look stupid
| and it irritates the pig.

(<http://minnie.cs.adfa.edu.au/cgi-bin/newsread?199044>)

Since e-mail is not as direct a communication medium as IRC, and since most academic Internet operators only tolerate relatively decent newsgroups on their net, one can - with a necessary amount of caution - predict that the language found on Usenet will be less informal and colloquial than on IRC (the difference between the examples (1) and (2) provisionally corroborates that assumption). Although we do not wish to make an a priori stratificational classification of the respective varieties of Dutch written in IRC and Usenet, it is safe to assume that the linguistic context in which IRC is produced is more informal than the context in which "Usenetese" emerges. Whereas, for one thing, IRC is immediate and online (hence the abundance of spelling and typing errors), the offline character of Usenet enables its users to reconsider and reread their contributions before posting them in the group (cf. Harrison 1998 for an overview of some linguistic characteristics of Usenet conversations).

Given then that IRC is likely to be more informal than Usenet, the synchronic hypothesis can be rephrased as follows: if the initial findings are replicated - to the extent that the new data confirm the larger distance between national and local language in Belgian Dutch -, the intermediate position Usenet appears to occupy on the informality dimension (which has the newspapers and IRC as its formal and informal pole), will manifest itself in a higher uniformity between the newspaper corpus and the Usenet corpus than between the newspaper corpus and the IRC corpus.

3.2. Methods and results

In the replication study, the same quantitative methodology was used as in the initial synchronic investigation. The dependent variable was uniformity in the naming of the 15 clothing concepts.⁴ The independent variables were *region* - Belgium vs. The Netherlands - and *stratification* - IRC vs. Usenet vs. newspapers.

The relevant findings can be found in the lower part of figure 1, which is organised in the same way as the top part. As can be observed in the figure, the new data unanimously confirm the synchronic hypotheses. As far as the original hypothesis is concerned, the larger distance between national language and local language in Belgian Dutch than in Netherlandic Dutch specifically manifests itself in IRC:

| | | |
|----------------|---|----------------|
| U'(B98,IRCB98) | < | U'(N98,IRCN98) |
| 82.63 | < | 89.82 |

In addition, the uniformity between newspapers and IRC is smaller in Netherlandic and (especially) Belgian Dutch than the uniformity between newspapers and Usenet:

U'(B98,IRCB98) < U'(B98,UsenetB98)
82.63 < 90.76

As a consequence, the formality cline “newspapers > Usenet > IRC” assumed in the refined synchronic hypothesis is also supported by the data.

4. Towards an explanation of the discrepancies between the original data and the replication

Although we now have two sets of findings which confirm the synchronic hypothesis, not all problems are solved yet. The remainder of this paper is devoted to two problematic discrepancies between the original data and the replication; the solutions we propose to these issues are currently being researched, or represent directions for further research.

Diachronically, the uniformity between the Belgian and Netherlandic 1998-newspaper samples is lower than the uniformity between the Belgian and Netherlandic 1990-newspapers ($U'(B98,N98) = 82.02 < U'(B90,N90) = 86.62$), which could suggest that the progressive convergence between Belgian and Netherlandic Dutch that was found in the original investigation has stagnated. This finding does not (or not yet), however, endanger our diachronic hypothesis, because the relatively lower lexical uniformity between the Belgian and the Netherlandic 1998-newspaper samples could well be due to *expertise variation* (cf. Geeraerts, Grondelaers & Bakema 1994: 112-115, 147 ff., 181 ff. and Grondelaers & Geeraerts 1998: 359 ff.): whereas most of the sources which constitute the corpus of the initial investigation were *fashion magazines*, the synchronic part of the replication study was based on general newspapers. If we assume that the Belgian specialists who write for fashion magazines are more aware of the Dutch standard than journalists working for general newspapers, then it is plausible that these specialists will sooner tend towards convergence.

The possible interference of expertise variation can be neutralised in two ways. First, we could differentiate for text type in the replication, viz. restrict the scope of the investigation to newspaper contributions which are exclusively devoted to fashion, and which are therefore more likely to have been written by authors with the same degree of expertise as the fashion magazine writers. A more obvious alternative, however, is to add a diachronic dimension to the 1998-data. Since the newspaper material from 1958 and 1978 (which is also available in the CONDIV-corpus), is comparable to the 1998-newspaper sample as far as the distribution of different text types is concerned, it is highly unlikely that a diachronic investigation which is based exclusively on the CONDIV-corpus should be corrupted by expertise variation.

Synchronically, the observation that the shopwindow material manifests markedly lower uniformity with the 1990-newspaper material than either of both types of Internet language with the 1998-newspaper sample ($U'(\text{shopB90,B90}) = 47.63 < U'(\text{IRCB98,B98}) = 82.63 < U'(\text{UsenetB98,B98}) = 90.76$), might – somewhat unintuitively – suggest that shopwindow language is less standardised than Internet conversations. However, shopwindow references and Internet conversations are so different linguistically that they cannot be compared on a methodologically sound basis: whereas the shopwindow materials are exclusively lexical, IRC manifests global informality and substandardisation which involves lexical as well as non-lexical parameters.

It is interesting to observe in this respect that the informality of IRC does not seem to be determined in the first place by lexical factors.⁵ No matter how idiom-prone and abusive IRC-users may be, the high uniformity between IRC and the newspapers suggest that chatters tend to be quite conformist when referring to everyday concepts such as clothing. The obvious question then is: what is the respective impact of lexical vs. non-lexical variables on the global level of IRC-informality? In order to answer that question, we are currently analysing the correlations between lexical, morphological, syntactic and phonetic factors which constitute the stratificational position of the respective source types. The principal aim of this new line of research is to compute an *informality index* per source or source type, on the basis of which we can subsequently outline a one or more-dimensional stratificational hierarchy.

4. Conclusion

Let us briefly recapitulate our findings. This paper reported on a series of usage-based investigations into the standardisation of Belgian Dutch. The first corpus study empirically confirmed the alleged diachronic convergence between Belgian and Netherlandic Dutch, as well as the larger synchronic distance between local and national language in Belgian Dutch (which is a major indication of the latter's delayed standardisation). Whereas the first investigation relied on shopwindow materials to represent local language, the second study (carried out to replicate the initial findings) made use of two types of Internet language, viz. Internet Relay Chat – the language of Internet chat sessions –, and the language of the Usenet platform. Although the initial synchronic findings were undeniably confirmed in the replication study, a number of discrepancies between the first and second study necessitate an additional phase in the research, in which the danger of unwanted expertise variation is neutralised, and in which the scope of the research is extended to include non-lexical factors. The latter in particular is a necessary prerequisite if we wish to unravel the stratificational complexity of Belgian Dutch; no matter how convincing, the present data do (not yet) permit such a disentanglement.

References

- Baetens Beardsmore, H. & P. Van de Craen (1979). 'The development of triglossia in Flemish Belgium'. In M. van de Velde & W. Vandeweghe (eds.), *Sprachstruktur, Individuum und Gesellschaft. Akten der 13. Linguistischen Kolloquiums 1*. Tübingen: Niemeyer.
- Bays, H. (1998). 'Framing and face in internet exchanges: a socio-cognitive approach'. *Linguistik Online* 98/1. Online publication on <http://viadrina.eu.v-frankfurt-o.de/~wjjournal/bays.html>
- Geeraerts, D., S. Grondelaers & P. Bakema (1994). *The structure of lexical variation. Meaning, naming, and context*. (Cognitive Linguistics Research 5). Berlin: Mouton de Gruyter.
- Geeraerts, D., S. Grondelaers & D. Speelman (1999). *Convergentie en divergentie in de Nederlandse woordenschat. Een onderzoek naar kleding- en voetbaltermen*. Amsterdam: Meertensinstituut.
- Geerts, G. (1975). *Voorlopers en varianten van het Nederlands. Een gedocumenteerd dia- en synchroon overzicht*. Leuven: Acco.
- Goossens, J. (2000). 'De toekomst van het Nederlands in Vlaanderen'. *Ons Erfdeel* 43, 3-13.
- Grondelaers, S. (2000). *De distributie van niet-anaforisch er buiten de eerste zinsplaats. Sociolexicologische, functionele en psycholinguïstische aspecten van er's status als presentatief signaal*. Doctoral dissertation K.U. Leuven.

- Grondelaers, S. & D. Geeraerts (1998). 'Vagueness as a euphemistic strategy'. In A. Athanasiadou & E. Tabakowska (eds.), *Speaking of emotions: conceptualization and expression*. Berlin: Mouton de Gruyter.
- Grondelaers, S., K. Deygers, H. Van Aken, V. Van Den Heede & D. Speelman (2000). 'Het CONDIV-corpus geschreven Nederlands'. To appear in *Nederlandse Taalkunde* 5/4.
- Harrison, S. (1998). 'E-mail discussions as conversation: moves and acts in a sample from a listserv discussion'. *Linguistik Online* 98/1. Online publication on <http://viadrina.euv-frankfurt-o.de/~wjournal/harrison.htm>
- Hentschel, E. (1998). 'Communication on IRC'. *Linguistik Online* 98/1. Online publication on <http://viadrina.euv-frankfurt-o.de/~wjournal/irc.html>
- Herring, S. (1996). *Computer-mediated communication: linguistic, social and cross-cultural perspectives*. Amsterdam: Benjamins.
- Ko, Kwang-Kyu (1996). 'Structural characteristics of computer-mediated language: a comparative analysis of Interchange discourse'. *Electronic Journal of Communication* 6/3. Online publication on <http://ublib.buffalo.edu/libraries/eresources/ejournals/records/ejc.html>
- Labov, W. (1972). 'Some principles of linguistic methodology'. *Language in Society* 1, 97-120.
- Paolillo, J. (1999). 'The virtual speech community: social network and language variation on IRC'. *Journal of Computer-Mediated Communication* 4/4. Online publication on <http://jcmc.huji.ac.il/paolillo.htm>.
- Velde, H. van de (1996). *Variatie en verandering in het gesproken Standaard-Nederlands (1935-1993)*. Doctoral dissertation K.U. Nijmegen.
- Willemyns, R. (1979). 'Norm en variatie in Vlaanderen'. In J. de Rooij (ed.), *Variatie en norm in de standaardtaal*. Amsterdam: Meertensinstituut.

Notes

1. It goes without saying that there are no shopwindow data for football concepts such as REFEREE, KICK-OFF or CORNER.
2. The new corpus was compiled by researchers from four Belgian and Netherlandic universities as a database for further investigation into issues related to the topic of this paper (cf. Grondelaers e.a.: 2000). The synchronic corpus contains 25.984.343 words of Belgian Dutch (IRC = 8.207.007 words; Usenet = 4.980.780 words; various newspapers = 12.796.556 words), and 19.505.008 words of Netherlandic Dutch (IRC = 6.965.291 words; Usenet = 7.748.436 words; various newspapers = 4.791.281 words). The diachronic corpus totals 5.666.763 words of Belgian and Netherlandic Dutch. Because it was purpose-built for the investigation of lexical convergence and divergence, the new corpus is habitually dubbed the "CONDIV-corpus".
3. In Computer Mediated Communication, the distinction "online vs. offline" equates the distinction between "synchronic" and "asynchronic". Whereas IRC is online and synchronic to the extent that it takes place in "real time" – in a text box in which users instantaneously respond to messages which appear on the screen -, Usenet is asynchronic and offline because there usually is a "propagation delay" between the composition of an e-mailmessage, the posting, the reception, and the eventual response (Paolillo 1999).
4. We restricted the synchronic replication to the concepts for which shopwindow data are available.
5. The limited impact of lexical factors on the global level of informality of a language variety has been observed in a.o. Goossens (2000: 9 ff.), who claims that the singularity of Belgian Dutch is determined to a larger extent by morpho-grammatical features (notably the form and/or declination of some of its determiners) than by lexical factors.