# SEL-HFDL — toelichting

## Het oude pakketje: SEL-HFDL.LZH (Ivo Gilisen, nov. 2011)

**Ivo Gilisen heeft dit pakket** in november 2011 gemaakt en verdeeld als `SEL-HFDL.LZH` (auteur destijds `ivo.gilisen@telenet.be`, nu `ivo.gilisen@pro-gen.be`). Het was een klein, op zichzelf staand installatiepakketje voor PRO-GEN, bedoeld als hulpmiddel bij het controleren van de **schrijfwijze van achternamen** in een bestand.

**Dit omvat** geen VTX-programma, maar **twee opgeslagen selectievoorwaarden** (telkens één regel, type SELECT) die bij installatie in de **personenlijst** (`pg-sel-1.cat`) terechtkomen:

1. *2de teken ACHTERNAAM is een hoofdletter*
   `ASC(MID$(Achternaam,2,1))>=65 AND ASC(MID$(Achternaam,2,1))<=90`
2. *2de teken ACHTERNAAM of PARTNER_ACHTERNAAM is een hoofdletter*
   dezelfde test, maar met `OR` ook op het veld `Partner_Achternaam` (dat enkel bestaat in bidprentjesbestanden — BIDPRENT.DEF, BIDPR.DEF, BIDP4IVO.DEF)

**Dit doet** het volgende: het **selecteert (of markeert)** alle personen waarvan het **tweede teken** van de achternaam een **hoofdletter (A–Z)** is. Zo spoor je namen op die (deels) in HOOFDLETTERS zijn ingevoerd — denk aan *MAes*, *SMETS* of *VANdijck* — typische invoeronnauwkeurigheden die je daarna uniform kunt rechtzetten. De tweede voorwaarde breidt die controle uit naar de **partner-achternaam**, handig in een bidprentjesbestand.

**Beperkingen** van deze aanpak:

1. **Enkel het tweede teken.** De test kijkt alleen naar positie 2 van de achternaam. Een verkeerde hoofdletter **verderop** in de naam wordt gemist — bv. *VanDijck*, *deBacKer*, *McCarThy*.
2. **Enkel onversierde A–Z (ASCII 65–90).** Een hoofdletter **mét accent** (É, È, Ë, Ç, Ö, Ä, Ü …) valt **buiten** 65–90 en wordt **niet** herkend. *GÉRARD* met É als tweede teken ontsnapt.
3. **Valse positieven bij legitieme hoofdletters.** Namen waar een hoofdletter op positie 2 **correct** is, worden onterecht opgepikt — vooral Nederlandse **IJ-namen** (*IJpelaar*, *IJsewijn*).
4. **Korte of lege achternaam.** Bij één teken (of een leeg veld) is er geen tweede teken; controle is niet mogelijk.
5. **Enkel het achternaam-veld.** Voornaam, voorvoegsel, alias en roepnaam worden **niet** gecontroleerd; en voorwaarde 2 dekt enkel `Partner_Achternaam`, niet `Partner2_/Partner3_/Gezel_Achternaam`.
6. **Veld-afhankelijkheid → foutmelding.** Voorwaarde 2 verwijst naar `Partner_Achternaam`, dat **alleen** in bidprentjesbestanden bestaat. In een gewoon bestand geeft PRO-GEN de melding *“niet-bestaand veld”* (gedocumenteerd in `Fout bij LZH-install.txt`).

7. **Detecteert, corrigeert niet.** Het pakket **selecteert of markeert** enkel; rechtzetten blijft handwerk.

## De oplossing — SEL-HFDL.VTX

Een nieuw, **DEF-bewust VTX-programma** (concept Ivo Gilisen, VTX-uitwerking Luc Klaps, 13-06-2026) bundelt de twee losse voorwaarden in één geheel. Het lost de *praktische* beperkingen op:

1. **Geen "niet-bestaand veld" meer** (beperking 6). De velden worden niet meer rechtstreeks benoemd, maar **tekst-gebaseerd** benaderd: het programma voegt een partner-/gezelveld enkel toe als `FIELDEXIST("Partner_Achternaam",1)=TRUE` (idem Partner2_/Partner3_/Gezel_) en leest de inhoud via `CONTENTS$(FIELDNUMBER(naam,1),1)`. Daardoor verschijnt er in **geen enkel bestandstype** nog een foutmelding.
2. **Volledige partner-/gezel-dekking** (deel van beperking 5). De VTX controleert — waar ze bestaan — **alle** partner- en gezelvelden, dus ook 2de/3de partners in een bidprentjesbestand.
3. **Één ingang i.p.v. twee regels** — één keuze werkt in elk bestandstype.
4. **Geen dubbeltellingen** — een persoon die op meerdere velden matcht, wordt via een `CHECKMARK`-buffer maar één keer geteld.
5. **Selecteren én markeren** vanuit de personenlijst (L of A), met statusscherm in BOX-kader en lopende teller.

De eerste VTX-versie (v1.00) nam bewust **dezelfde A–Z-test** over (identiek resultaat aan Ivo's origineel). Versie **v1.01** lost daarbovenop de inhoudelijke beperkingen **1, 2, 3, 4 en 5-deels** op (zie "Wat we bouwen"); enkel het **corrigeren** (7) blijft per definitie handwerk.

## Kunnen die overige beperkingen weg? (alles behalve 7)

- **(1) Enkel het tweede teken → ja.** Loop met een `FOR`-lus over álle tekens vanaf positie 2 tot `LEN(naam)` en test elk teken; dan vang je ook *VanDijck*, *deBacKer*.
- **(2) Accenthoofdletters → ja, en properder.** Vervang `ASC(...)>=65 AND <=90` door een vergelijking met de kleine-letter-versie: een teken is een hoofdletter als `MID$(naam,i,1) <> LCASE$(MID$(naam,i,1))`. Dan tellen É/È/Ë/Ç/Ö/Ä/Ü mee. **Voorbehoud (te verifiëren):** dit werkt enkel als PRO-GEN's `LCASE$` die accenttekens effectief omzet — eerst nakijken.
- **(3) IJ-valse positieven → slechts deels.** De duidelijke **IJ**-regel kun je expliciet uitsluiten (teken 1 = I en teken 2 = J → niet markeren). Het algemene "bedoeld vs. fout" (Mc-, Mac-, De-…) blijft een oordeel dat je niet volledig kunt automatiseren.
- **(4) Korte/lege achternaam → niets nodig.** Bij één teken valt er niets te controleren; "geen treffer" is correct. Past (1) toe, dan lost dit zich vanzelf op.
- **(5-rest) Voornaam/voorvoegsel/alias → ja.** Voeg die velden toe aan de te-controleren lijst (dezelfde `FIELDEXIST`/`FIELDNUMBER`-aanpak). Let op: een

voorvoegsel als *van*/*de* is normaal klein, dus daar kan een hoofdletter net gewenst óf fout zijn.

- **(7) Detecteert, corrigeert niet → kan niet.** Een selectie-/markeer-VTX schrijft niets terug naar de database; rechtzetten blijft per definitie handwerk.

## Het geval Van der Elst — de woordgrens-regel

Dit testgeval legt een zwakte in voorstel (1) bloot. Het hangt af van hoe de naam is opgeslagen:

- **"van der" in het Voorvoegsel-veld, Achternaam = "Elst":** geen probleem — teken 2 van *Elst* = l (klein), geen treffer, ook niet bij een volledige scan.
- **Heel "Van der Elst" in het Achternaam-veld:** de oude teken-2-test kijkt naar de a van *Van* → toevallig geen treffer. Maar een **volledige scan** (1) stoot verderop op de hoofdletter **E** van *Elst* → en zou *Van der Elst* **onterecht markeren**. Een nieuwe valse positief.

**Voorstel tot oplossing:** markeer een hoofdletter op positie i alleen als het **vorige teken óók een letter** is (hoofdletter *midden in een woord*), níet als ze volgt op een **spatie, koppelteken of apostrof** (begin van een nieuw woord):

| naam | gedrag |
|---|---|
| *Van der Elst* | E volgt op spatie → niet gemarkeerd ✅ |
| *VanDerElst (aaneen, fout)* | D/E volgen op een letter → wel gemarkeerd ✅ |
| *MAes* | A volgt op letter → wel gemarkeerd ✅ |
| *Janssens-Peeters* | P volgt op koppelteken → niet gemarkeerd ✅ |
| *O'Brien* | B volgt op apostrof → niet gemarkeerd ✅ |

**Keerzijde:** met die regel wordt een verkeerd gekapitaliseerd **voorvoegselwoord** — bv. *Van Der Elst* (de D van *Der* volgt op een spatie) — net **niet** meer opgemerkt. Dat is opnieuw het "bedoeld vs. fout"-oordeel dat niet volledig automatiseerbaar is. En de **IJ-namen** overleven deze regel óók (de J volgt op de letter I), dus die blijven een aparte uitzondering vragen.

## Uitbreidingsvoorstellen

Samengevat, in oplopende moeilijkheid:

1. **Volledige naamscan** i.p.v. enkel teken 2 — `FOR`-lus over alle tekens (lost 1 op).
2. **Woordgrens-regel** — enkel een hoofdletter na een letter markeren, niet na spatie/koppelteken/apostrof (maakt 1 bruikbaar voor samengestelde namen als Van der Elst).
3. **LCASE$-test i.p.v. ASC-bereik** — om ook accenthoofdletters te vangen (lost 2 op — mits `LCASE$`-accentcheck geverifiëerd).
4. **IJ-uitzondering** — expliciete regel tegen valse positieven bij Nederlandse IJ-namen (lost 3 deels op).

5. **Extra velden** — Voornaam, Voorvoegsel, Alias, Roepnaam toevoegen aan de controle, elk eventueel met een eigen casing-regel (lost 5-rest op).

Niet automatiseerbaar blijven: het algemene onderscheid **bedoelde vs. foute** hoofdletter (3, deels) en het **corrigeren** zelf (7).

## Wat we bouwen — versie v1.01

**Versie-afspraak.** Ivo's origineel was een opgeslagen selectievoorwaarde **zonder versienummer** (*"Selectie 2de teken hoofdletter", z/n*). De VTX-herwerking van Luc Klaps start op **v1.00** (13-06-2026) en is met **v1.01** (16-06-2026) uitgebreid tot het onderstaande.

**v1.01 omvat:**

- **Volledige naamscan** i.p.v. enkel teken 2 — vangt ook hoofdletters verderop in de naam (*VanDijck*, *deBacKer*) (beperking 1).
- **Woordgrens-regel** — een hoofdletter wordt enkel gemeld als ze ná een letter staat, niet na een spatie, koppelteken of apostrof; zo blijft *Van der Elst* terecht onaangeroerd.
- **Voorvoegsel-regel** — in het voorvoegsel mag enkel de 1e letter een hoofdletter zijn: *Van Den …* wordt gemeld, *Van den …* niet. De voorvoegsel-grens komt uit PRO-GENs eigen voorvoegsellijst (`PRO-GEN.PFX`), dus ook als de voorvoegsel-herkenning uit staat. Samengestelde namen met een echte basisnaam (*de Berghes van Grimbergen*) blijven ongemoeid.
- **Accent-hoofdletters** (É, Ä, Å, Ç, Æ, Ö, Ü, Ñ uit PRO-GENs tekentabel) tellen mee (beperking 2). Detectie via ASC-codes — PRO-GEN vergelijkt tekst hoofdletter-ongevoelig, dus een `LCASE$`-vergelijking werkt hier net niet.
- **IJ-uitzondering** — Nederlandse IJ-namen (*IJpelaar*) worden niet gemarkeerd (beperking 3).
- **Mc/Mac-uitzondering** — *McDonald*, *MacArthur* worden niet gemeld.
- **Romeinse-cijfer-uitzondering** — adellijke nummering (*Karel V*, *Rieneck II*, II t/m XL) wordt niet gemeld.
- **Achternaam-velden** — Achternaam plus Partner_/Partner2_/Partner3_/Gezel_Achternaam wáár ze bestaan, DEF-bewust via `FIELDEXIST`.
- **Selecteren én markeren** (L of A), BOX-statusscherm met teller, en `SET OPTIONS` wordt bewaard en hersteld.

**Blijft bewust buiten beschouwing:** Voornaam, Alias en Roepnaam (een hoofdletter midden in een voornaam komt nauwelijks voor) en het **corrigeren** zelf (beperking 7) — een selectie-/markeer-VTX schrijft niets terug naar de database; rechtzetten blijft handwerk.

## In de nieuwsbrieven

Aangekondigd in de PGggLIMBURG-nieuwsbrief van **24-11-2011** (samen met SELVNAAM), met downloadlink `sel-hfdl.lzh`.

## Auteurs

Ivo Gilisen (nieuwsbrief 24-11-2011) • Luc Klaps v1.01 (gereviseerd 2026) • Johan Mulderij & D.J. Scholte in 't Hoff.