

Artländer Phonetik

Ein regionales phonetisches Algorithmus

Status: Working Paper | Version 1.0

Publikationsdatum: Sep 2026

Autor: Ronald Voortman

Lizenz: Dieses Werk ist lizenziert unter einer [Creative Commons Namensnennung-Nicht kommerziell 4.0 International Lizenz](#).



Zusammenfassung

Dieses Whitepaper präsentiert die Entwicklung und Validierung der Artländer Phonetik, eines spezialisierten phonetischen Algorithmus, der entwickelt wurde, um die extreme Schreibvariantenvielfalt in historischen Kirchenbüchern der Region Artland (Badbergen, Menslage und Quakenbrück) zwischen 1600 und 1800 effektiv zu strukturieren. Eine auf der Kölner Phonetik basierende Nullmessung zeigte, dass generische, moderne Algorithmen bei niedersächsischen Regionaldaten versagen: Sie erzielen lediglich 8% vollständige Namensübereinstimmungen bei einem problematischen Rauschen von 36% (falsche Verknüpfungen zwischen nicht verwandten Familien). Indem die Architektur auf einer geschichteten Analyse amtlicher Orthographie, den sprachlichen Gesetzmäßigkeiten aus dem Werk von Hermann Vehslage (1908) und einem fünfstufigen Reduktionsprozess unter Beibehaltung des Anlauts basiert, erzielt die Artländer Phonetik eine fundamentale Verbesserung. Anhand eines Korpus von 175.270 Quellennachweisen steigt der Prozentsatz vollständiger Darstellungen auf 60%, während das Rauschen auf vernachlässigbare 8% sinkt. Damit bildet dieser Algorithmus das zuverlässige Fundament für das Artland Record Matching System (Artland-RMS) und automatisierte genealogische Rekonstruktionen.

Einleitung

Wer sich in die Genealogie und demografische Geschichte des Artlandes vertieft, stößt rasch auf ein hartnäckiges historisches Hindernis: die enorme Willkür in der Schreibweise von Familiennamen in den frühneuzeitlichen Kirchenbüchern. Dieselbe Familie konnte innerhalb weniger Jahrzehnte von verschiedenen Geistlichen als Möhlmann, Molman, Molenhus oder zur Mohlen notiert werden. Diese Varianz ist kein Zeichen administrativer Nachlässigkeit, sondern das direkte Resultat eines permanenten soziallyinguistischen Spannungsfelds. Einerseits gab es die gesprochene Mundart der lokalen Bevölkerung (das niedersächsische Plattdeutsch); andererseits existierte die amtliche Schriftsprache von auswärtigen Geistlichen, die versuchten, die Namen in hochdeutscher oder lateinischer Orthographie zu erfassen.

Für die moderne historische und genealogische Forschung – bei der große Datenbanken wie die Personendatenbank von vortmes.nl Tausende von Datensätzen miteinander verknüpfen müssen – stellt diese unstrukturierte Namensvarianz einen ernsthaften Engpass dar. Standardmäßige phonetische Algorithmen wie die Kölner Phonetik, die für die moderne deutsche Sprache optimiert sind, versagen in diesem regionalen Kontext gravierend. Ihnen fehlen regionale Lautgesetze, sie wenden zu strenge Reduktionen an oder erzeugen in inakzeptablem Maße Verunreinigungen, indem sie nicht verwandte Familien miteinander verknüpfen.

Dieses Whitepaper beschreibt die methodologische und technische Notwendigkeit hinter der Artländer Phonetik. Anhand eines großflächigen Datensatzes von über 175.000 Kirchspielsdaten wird dargelegt, wie das Wechselspiel zwischen amtlicher Orthographie und regionalgebundenen Lautgesetzen zu einem neuen, geschichteten Algorithmus führte. Indem die sprachwissenschaftlichen Erkenntnisse von Dialektologen wie Hermann Vehslage in einem fünfstufigen Reduktionsprozess – einschließlich der bewussten Sicherung des Anlauts zur Wahrung des Familienkerns – verankert werden, schlägt die Artländer Phonetik die Brücke zwischen frühneuzeitlicher Regionalgeschichte und Datentechnologie des einundzwanzigsten Jahrhunderts. Die Ergebnisse zeigen, dass dieser regionale Algorithmus nicht nur die Datenqualität drastisch bereinigt, sondern auch den Weg für eine automatisierte Record Linkage im historischen Maßstab ebnet.

2. Datensatz und linguistischer Kontext

Die Entwicklung und Validierung der Artländer Phonetik basieren auf einem großflächigen, historischen Datensatz aus der Personendatenbank von vortmes.nl. Dieser Datensatz umfasst kirchliche Quellen aus der Region Artland, womit ein repräsentatives Fundament zur Analyse historischer Namensvariationen und Schreibfluktuationen zwischen 1600 und 1800 geschaffen wurde.

2.1 Datenauswahl und Quellen

Für die Zusammenstellung des Korpus wurden fundierte Kirchenbücher aus drei Kerngebieten herangezogen:

- **Badbergen:** Tauf-, Heirats- und Sterberegister für den Zeitraum 1671–1799.
- **Menslage:** Tauf-, Heirats- und Sterberegister für den Zeitraum 1694–1799.
- **Quakenbrück:** Ausschließlich das Sterberegister für den Zeitraum 1667–1799. Da Quakenbrück eine heterogenere städtische Bevölkerung aufwies, die keineswegs immer autochthon aus dem Artland stammte, wurde hier bewusst eine Beschränkung auf die Sterbebücher gewählt, um eine Verzerrung der Migrationsdynamik zu minimieren.

Der initiale Datensatz umfasste insgesamt 204.884 Quellennachweise. Nach einer strengen Datenbereinigung (Data-Cleaning), bei der unvollständige, unleserliche oder fehlende Familiennamen herausgefiltert wurden, verblieb eine Quellengruppe von 175.270 Einträgen.

2.2 Namensvarianz und Normalisierung

Innerhalb des bereinigten Datensatzes von 175.270 Einträgen wurden insgesamt 2.635 einzigartige Schreibvarianten von Familiennamen identifiziert. Dies bedeutet, dass eine individuelle Schreibvarianten im Durchschnitt 66,5 Mal vorkommt.

Mittels historisch-onomastischer Analyse und Normalisierung wurde diese große Vielfalt historischer Schreibweisen auf 700 einzigartige, standardisierte Familiennamen reduziert. Dies führt innerhalb des Datensatzes zu den folgenden statistischen Verhältnissen:

- Ein standardisierter Familienname weist durchschnittlich 3,76 verschiedene Schreibvarianten auf.
- Jede einzigartige Schreibvariante kommt in den Quellen durchschnittlich 66 Mal vor.

- Jeder standardisierte Familienname repräsentiert im Durchschnitt 250 Einträge (über die Varianten verrechnet).

Anmerkung: Diese Zahlen betreffen Mittelwerte; in der Praxis weist die Streuung je Familienname eine natürliche Varianz auf. Zudem impliziert ein geteilter Familienname nicht automatisch eine einzige genealogische Stammfamilie; so kommt der Familienname Meijer beispielsweise in mehreren, voneinander unabhängigen Familienlinien innerhalb des Artlandes vor.

2.3 Hochdeutsche und niederdeutsche Einflüsse durch auswärtige Schreiber

Die enorme Vielfalt an Schreibvarianten ist kein Zufall, sondern das direkte Resultat des sozilinguistischen Kontexts, in dem die Urkunden verfasst wurden. Viele der Kirchenbücher und administrativen Akten wurden von Geistlichen, Küstern oder Verwaltungsbeamten geführt, die keineswegs immer aus dem Artland selbst stammten. Dies führte zu einer permanenten Spannung und Vermischung zweier Sprachsysteme:

- **Das Niederdeutsche (Plattdeutsch):** Die gesprochene Regionalsprache der lokalen Bevölkerung, in der Klänge und phonetische Gesetzmäßigkeiten für die Region typisch sind. Namen wurden häufig phonetisch so aufgeschrieben, wie sie im Mund des Sprechers klangen (beispielsweise die Wiedergabe von Langvokalen, die Verhaertung von Konsonanten oder spezifische Ausgänge wie -mann versus -man).
- **Das Hochdeutsche (Hochdeutsch):** Die formale amtliche oder kirchliche Schriftsprache, die von auswärtigen Schreibern oktroyiert wurde. Sie versuchten, die lokalen, niederdeutschen Namen in hochdeutsche Rechtschreibkonventionen zu fassen.

Dies führte dazu, dass derselbe Familienname von unterschiedlichen Schreibern (oder sogar von demselben Schreiber zu verschiedenen Zeitpunkten) völlig unterschiedlich geschrieben werden konnte – reichend von direkten niederdeutschen phonetischen Transkriptionen bis hin zu stark hochdeutsch adaptierten oder lateinischen Schreibweisen in den offiziellen Registern.

3. Phonetische Interferenzen und amtliche Orthographie

Die Herkunft eines Geistlichen übte im 18. Jahrhundert einen ausschlaggebenden Einfluss auf die Orthographie in den Tauf-, Heirats- und Sterberegistern (Kirchenbücher) im Artland aus. Da keine standardisierte Rechtschreibung existierte, fungierte der Pfarrer oder Pastor als der eigentliche Lexikograf seiner Gemeinde. Seine Feder wurde durch eine Kombination aus drei Kräften gesteuert: seiner eigenen regionalen Herkunft (phonetische Interferenzen), seinem Bildungsniveau (lateinische Tradition) und dem Grad, in dem er die lokale niedersächsische Volkssprache (Mundart) beherrschte.

3.1 Geografische Mobilität und Herkunft (Allochthone Schreibstile)

Geistliche stammten in dieser Periode oft von außerhalb des direkten Dorfes oder sogar von außerhalb der Region. Lutheraner studierten an Universitäten in Jena, Helmstedt oder Rinteln; Reformierte zog es bisweilen nach Groningen oder Franeker; katholische Priester wurden in bischöflichen Priesterseminaren wie Münster oder Osnabrück ausgebildet.

Wurde ein Prediger oder Pastor aus einer anderen Region berufen, brachte er seine eigenen phonetischen Schreibreflexe mit:

- **Vokalwechsel:** Ein Pastor aus dem Osnabrückschen schrieb die kurzen und langen Vokale anders als ein Kollege aus der Grafschaft Tecklenburg oder aus Ostfriesland. Der Laut, der lokal wie ein offenes ä oder ö klang, wurde von einem auswärtigen Geistlichen häufig nach dem notiert, was sein eigenes Ohr zu hören glaubte (beispielsweise die Vertauschung zwischen e und ä oder o und u).
- **Konsonantenverschiebung:** Regionale Variationen bei der Wiedergabe scharfer und weicher Konsonanten (wie t und d oder der scharfen s- und sch-Laute) schlugen sich direkt im Familiennamen nieder. Ein Nachname, der lokal auf einen harten Plosiv endete, konnte unter der Feder eines auswärtigen Predigers plötzlich eine weiche Endung erhalten oder umgekehrt.

3.2 Der lateinisch-humanistische Filter

In römisch-katholischen Priestereien (und zu Beginn des 18. Jahrhunderts noch stark beim reformierten Klerus) war das Latein die amtliche Geschäftssprache für die

Urkunden. Dies führte zu einer kuriosen doppelten Entstellung autochthoner artländischer Namen:

- **Latinisierung von Patronymen und vor allem Familiennamen:** Personen erhielten lateinische Namensuffixe aufgepfropft. Ein Bauernsohn Hermann wurde plötzlich Hermannus, dessen Vater Gerhard wurde zu Gerhardus. Der Nachname wurde bisweilen phonetisch latinisiert oder mit Genitivendungen versehen (wie Henricus filius Joannis).
- **Die Rückkehr zur Volkssprache:** Viele Pastoren und Pfarrer handhabten ein Mischmasch: Die feste Struktur war lateinisch (baptizatus est, matrimonio junxi), aber die Eigennamen selbst mussten aufgeschrieben werden. Dies zwang den Geistlichen dazu, eine phonetische Übersetzung aus dem Niedersächsischen ins lateinische Alphabet vorzunehmen.

3.3 Die Kluft zwischen hochdeutschem Amtssidiom und artländischer Mundart

Der Geistliche schrieb in der Regel eine Form des (Früh-)Hochdeutschen oder der Amtssprache Latein, während die Bevölkerung ausschließlich Niederdeutsch (Westfälisch/Artländer Platt) sprach. Dies erzeugte ständige Reibungen auf dem Papier:

- **Hyperkorrektur:** Ein Pastor, der bemerkte, dass die lokale Bevölkerung einen bestimmten Laut aussprach, versuchte diesen bisweilen zu einer offizielleren hochdeutschen Schreibweise zu „bereinigen“, was zu Schreibweisen führte, die im Dorf selbst nie so ausgesprochen wurden.
- **Lautnachahmung (Ad-hoc-Phonetik):** Wenn ein Geistlicher weniger belesen war oder einer festen Orthographie weniger Wert beimaß, schrieb er rein nach Gehör (phonetic spelling). Dadurch konnte derselbe Familienname innerhalb eines Jahrzehnts vom selben Pastor auf drei verschiedene Arten geschrieben werden, je nachdem, wie der Bauer oder die Magd ihren Namen während der Registrierung aussprach.

[Folgen für die genealogische Forschung im Artland](#)

Die Kombination aus Herkunft, lateinischer Tradition und Mundart führt dazu, dass Nachnamen in den Kirchenbüchern des Artlandes aus dem 18. Jahrhundert selten statisch sind. Derselbe Familienname kann in Menslage anders zu Papier gebracht worden sein als in Badbergen oder Quakenbrück, schlicht weil der Prediger dort aus einer anderen Region stammte oder eine andere Ausbildung genossen hatte. Dies erklärt, warum Namensvarianten in dieser Region kein Zeichen administrativer

Schlamperei an sich sind, sondern die direkte Spur der kulturellen und geografischen Biografie des schreibenden Geistlichen bilden.

4. Der Beitrag von Hermann Vehslage (1908): Historisch-lautgesetzliche Verankerung

Wo sich die vorherigen Kapitel primär auf die quellenkundliche Einbettung und den Einfluss der amtlichen Orthographie konzentrierten, bietet die Dissertation von Hermann Vehslage aus dem Jahr 1908 (Die Mundart der Badberger Gegend) eine unersetzliche sprachwissenschaftliche und dialektgeografische Fundierung. Für die Entwicklung und Verfeinerung der Artländer Phonetik fungiert Vehslages Werk nicht als statische Wortliste, sondern als methodologischer Kompass, der erklärt, warum bestimmte Lautmutationen und Schreibvarianten in den historischen Quellen auftreten.

4.1 Einblick in die niederdeutschen Lautgesetze (Vokalismus und Konsonantismus)

Die Transkriptionen in den Registern des 17. und 18. Jahrhunderts sind keine willkürlichen Schreibfehler, sondern spiegeln das permanente Wechselspiel zwischen der gesprochenen Mundart (Plattdeutsch) und der hochdeutschen

Amtssprache der Schreiber wider. Vehslages detaillierte Analysen der Lautgesetze legen spezifische Muster offen, die direkt in die Architektur des Artland-Algorithmus eingearbeitet wurden:

- **Nebensilben und Reduktion:** Die starke Abschleifung unbetonter Silben erklärt, warum Endungen in Tauf- und Heiratsbüchern stark fluktuieren können. Suffixe wie -ing (lynigk), -scap oder die Schwächung von Präfixen (gi-, bi-) zu reduzierten Vokalen oder vollständigen Elisionen (wie maux oder andmm) rechtfertigen die strenge Auslaut-Normalisierung in Phase 4 des Algorithmus.
- **Konsonantenverschiebungen und Assimilationen:** Vehslage beschreibt, wie inlautende Kombinationen (wie mb, mn, gg, ld, nd, rd) und auslautende



Abbildung 1. Hermann Vehslage (1883–1917) als Freiwilliger bei der Feldartillerie (um 1915). Der Autor der wegweisenden Dissertation Die Mundart der Badberger Gegend (1908) trat im Januar 1915 als Kriegsfreiwilliger in den Militärdienst ein und fiel im September 1917 an der rumänischen Front.

Verhärtungen (d zu t, g zu x) strukturell auftreten. Der Algorithmus nutzt dieses Wissen, um in Phase 2 und 3 gezielte Transformationen anzuwenden, damit historisch verwandte Stämme nicht zu Unrecht durch allgemeine Reduktionsregeln verunstaltet werden.

- **Die Rolle der Aussprachevarianten:** Das Phänomen, bei dem sich der stimmhafte oder stimmlose Charakter von Konsonanten verschiebt (wie die Neigung zu scharfen s-Lauten oder die spezifische Behandlung der Halbvokale w und j), bildet die Rationale hinter der Sibilanten-Harmonisierung und dem flexiblen Umgang mit Konsonantenclustern in den frühen Phasen der Artländer Phonetik.

4.2 Dialektgeografie und die geografische Grenze als Kontext

Im Schlussteil seiner Studie (insbesondere § 72) platziert Vehslage das Artland in einer breiteren geografischen Perspektive. Er zeigt auf, dass die Region sprachlich und historisch nicht einfach unter den Sammelbegriff des westfälischen Hauptdialekts gefasst werden kann, sondern einzigartige Merkmale aufweist, die auf die frühmittelalterliche Rodungsgeschichte und Migration aus dem Emsland und dem Niederstift Münster zurückzuführen sind.

- **Bedeutung für Record Matching:** Diese historische Grenzdynamik erklärt, warum generische Algorithmen versagen. Die Bevölkerung des Artlands befand sich in einer Übergangszone, in der Einflüsse von außen (unter anderem durch die Kultivierung seitens des Klosters Corvey und den Handel rund um Quakenbrück und Wulften) für eine einzigartige Palette an Namensvarianten sorgten.
- **Validierung der Whitelists:** Vehslages Feststellung, dass sich die Mundart auf spezifische Weise zu ihren nördlichen und östlichen Nachbarregionen verhält, validiert die Notwendigkeit der in Phase 3 verwendeten Whitelists und stammesspezifischen Korrekturen innerhalb der Artländer Phonetik. Dies verhindert, dass regional gebundene Familiennamensvarianten durch rigide, moderne Regeln weggestrichen werden.

Durch die Verknüpfung der empirischen Erkenntnisse aus Vehslages Dissertation mit der Datenbankstruktur von vortmes.nl überbrückt die Artländer Phonetik die Kluft zwischen theoretischer Dialektologie des 19. Jahrhunderts und Computational Genealogy des 21. Jahrhunderts.

5. Die Nullmessung: Die Kölner Phonetik als Referenz

Um die Notwendigkeit und Effektivität eines spezifisch regionalen phonetischen Algorithmus nachzuweisen, wurde der Datensatz einer Nullmessung auf Basis der Kölner Phonetik unterzogen – dem traditionellen deutschen Standard für den phonetischen Namensabgleich.

5.1 Ergebnisse der Nullmessung

Die Ergebnisse dieser Messung belegen fundamentale Einschränkungen bei der Anwendung auf frühneuzeitliche Regionaldaten:

- **8% vollständige Wiedergabe:** Lediglich acht Prozent der Familiennamen ergaben einen 100%-Match, bei dem alle Schreibvarianten korrekt und ohne Hinzufügung von Rauschen verknüpft wurden.
- **56% unvollständige Wiedergabe:** Mehr als die Hälfte der Ergebnisse war unvollständig; das Algorithmus übergang relevante Varianten, da die starren Kölner Regeln für niedersächsische und regionale Lautmutationen zu streng oder ungeeignet sind.
- **36% Rauschen:** Gut ein Drittel der Ergebnisse enthielt Verunreinigungen in Form zu Unrecht verknüpfter Familiennamen.

5.2 Was ist „Rauschen“ beim phonetischen Abgleich?

Wenn ein generisches Algorithmus wie die Kölner Phonetik auf diesen historischen Datensatz losgelassen wird, entsteht ein beträchtlicher Prozentsatz an Rauschen (ermittelt bei 36%). Im Kontext historischer und genealogischer Datenbanken ist unter Rauschen folgendes zu verstehen:

Rauschen tritt auf, wenn ein phonetisches Algorithmus unterschiedliche, distinkte Familiennamen fälschlicherweise miteinander verknüpft, weil ihre Klangstruktur innerhalb des Algorithmus zufällig einander angenähert wird.

Ein treffendes Beispiel aus der Analyse hierzu ist, wie die Kölner Phonetik die Familiennamen *Elting* und *Oldenhage* einander gleichsetzt. Für ein Algorithmus, das für die moderne deutsche Sprache und Lautreduktion konzipiert ist, mag dies mathematisch

vertretbar erscheinen, genealogisch und historisch ist es jedoch völlig inkorrekt; in der Realität handelt es sich um zwei völlig verschiedene Familien und Entitäten im Artland.

Wo ein regional ausgerichteter Algorithmus die Unterscheidung auf Basis regionaler Phonetik und Kontext wahrt, greift ein starres Algorithmus durch dieses fehlgeleitete Gruppierungsverhalten völlig daneben. Solche False Positives verunreinigen die Suchergebnisse, wodurch Forschern Listen präsentiert werden, in denen nicht verwandte Familien miteinander vermischt werden.

Diese Mängel und der hohe Prozentsatz an Rauschen unterstreichen, warum die Artländer Phonetik geboten war: ein Algorithmus, das spezifisch auf die regionalen Lautgesetzmäßigkeiten abgestimmt ist, um die Suchergebnisse drastisch zu bereinigen und zu verbessern.

6. Die Architektur und Funktionsweise der Artländer Phonetik

Wo generische Algorithmen wie die Kölner Phonetik an regionalen und historischen Schreibweisen scheitern, verfolgt die Artländer Phonetik einen geschichteten Ansatz. Der Algorithmus ist speziell dafür konzipiert, die Kluft zwischen den wechselnden Transkriptionen auswärtiger Schreiber und der zugrundeliegenden genealogischen Realität im Artland zu überbrücken.

Anstatt einer blinden Lautreduktion wendet der Algorithmus eine Abfolge von fünf funktionalen Phasen an. Dadurch werden grafische und sprachliche Störeinflüsse schrittweise bereinigt, bevor die eigentliche phonetische Kodierung stattfindet.

6.1 Phase 1: Prä-Processing und Normalisierung der Grafie

Bevor Laute analysiert werden, werden typografische und regionale Notationsunterschiede, die nichts mit der tatsächlichen Aussprache zu tun haben, begradigt:

- **Grafische Uniformierung:** Die Umwandlung aller Zeichen in Kleinbuchstaben sowie die Normalisierung von Umlauten ($\ddot{a} \rightarrow a$, $\ddot{o} \rightarrow o$, etc.) und Doppelvokalen ($aa \rightarrow a$).
- **Standardisierung wechselnder Buchstaben:** Häufige frühneuzeitliche Verschreibungen werden direkt eindeutig gemacht, wie die Ersetzung von v durch f , dt durch t und diverser y -Varianten durch i .
- **Sibilanten-Harmonisierung:** S-Laute, die auf unterschiedliche Weise notiert wurden (wie sz , β und z), werden auf eine einheitliche Basis zurückgeführt.

6.2 Phase 2: Regionale Lautkorrekturen und niedersächsische Besonderheiten

Die Urkunden zeigen ein permanentes Wechselspiel zwischen der lokalen gesprochenen Mundart (Plattdeutsch) und dem hochdeutschen Amtsstil des Schreibers. Diese Phase fängt diese spezifischen Lautmutationen ab:

- **Stamm- und Lautkorrekturen:** Spezifische regionale Suffixe und Verballhornungen (wie $-breh \rightarrow -brede$ oder $-ultz \rightarrow -ult$) werden abgefangen.

- **Kontextabhängige qu-Handhabung:** Der Algorithmus bestimmt anhand der umliegenden Buchstaben, ob ein *qu* phonetisch als Konsonant (*k*) oder als Lautgruppe (*kw*) behandelt werden muss.

6.3 Phase 3: Cluster-Sicherung und stammspezifische Korrekturen

Nicht jeder Laut darf beliebig verkürzt oder angepasst werden; bestimmte Buchstabenkombinationen tragen die historische Identität eines Familiennamens.

- **Längungs-h-Schutz:** Die Beibehaltung des Längungs-h in spezifisch geschützten Kontexten (wie *-horn* oder *-haus*), während es anderswo (beispielsweise *behl* / *beel* oder *mohr* / *moor*) sehr wohl normalisiert wird.
- **Whitelists für Ausnahmestämme:** Wo nötig, weicht der Algorithmus kontrolliert von den Standardregeln für spezifische Familien ab. Man denke hierbei an gezielte Transformationen von *-nn-* zu *-nd-* oder von *-ll-* zu *-ld-* für bekannte Stämme (wie *hillebrandt* → *hildebrandt* oder *brunnert* → *brundert*).
- **Inlaut-Optimierung:** Das Auffüllen oder Wiederherstellen fehlender Vokale in Konsonantenclustern zur Gewährleistung der Wiedererkennbarkeit des Stammes.

6.4 Phase 4: Auslaut-Normalisierung (Endungs-Harmonisierung)

Die meisten Varianten in historischen Registern entstehen am Ende eines Familiennamens (im Auslaut). Diese Phase bringt dort Struktur hinein:

- **Schutz fester Endungen:** Historische Stämme, die intakt bleiben müssen (wie *-horst*), werden sichergestellt.
- **Namensuffix-Normalisierung:** Administrative Entstellungen wie Genitivendungen (*-des*) oder verballhornte Endungen (*-ing*) werden auf den korrekten Familiennamenstamm korrigiert (einschließlich gezielter Konversionen wie *rossing* → *rossmann*).
- **Endklangkorrektur:** Der einheitliche Abschluss von Wörtern, wie die Umwandlung eines End-*d* in ein *-t* oder das Hinzufügen eines stabilisierenden *-en*.

6.5 Phase 5: Definitive phonetische Reduktion

Erst wenn der Name vollständig normalisiert und von historisch-grafischen Verunreinigungen befreit ist, wird der definitive phonetische Schlüssel generiert. Dies geschieht über einen verschärften Reduktionsalgorithmus, der auf den Prinzipien der klassischen Phonetik beruht, nun aber mit den bereinigten, harmonisierten Artland-

Daten gespeist wird. Dadurch werden irrelevante Schreibvarianten fehlerfrei unter denselben Schlüssel subsummiert, während destruktives Rauschen (wie die fehlerhafte Verknüpfung zwischen Elting und Oldenhage) effektiv verhindert wird.

Beim klassischen Kölner-Phonetik-Algorithmus wird der Anlaut – abgesehen von einigen spezifischen Regeln – auf Basis derselben Zifferntabelle wie der Rest des Wortes kodiert. So erhält ein Name, der mit einem C beginnt, beispielsweise eine 8 (oder eine 4 für einen vokalähnlichen Anlaut in bestimmten Positionen), wodurch Wörter mit einem vergleichbaren Anfangsklaut direkt unter denselben Hauptcode fallen.

Bei der Artländer Phonetik wurde bewusst beschlossen, den normalisierten Anlaut vollständig intakt zu lassen. Die Argumente hierfür sind historischer und genealogischer Natur:

- **Wahrnehmung des Familienkerns:** In historischen Urkunden und Regionalregistern (wie jenen im Artland) ist der Anlaut eines Familien- oder Hofnamens äußerst stabil und Träger der einzigartigen Identität. Die Berücksichtigung des Anlauts in einem Zifferncode würde zu unerwünschten „falschen“ Treffern zwischen klanglich verwandten, aber genealogisch völlig unabhängigen Geschlechtern führen.
- **Strikte Trennung der Phasen:** Da die phonetische Reduktion (Phase 5) erst stattfindet, nachdem alle grafischen Störeinflüsse in den vorherigen Phasen eliminiert wurden, ist eine Kodierung des Anlauts überflüssig geworden. Der normalisierte Anlaut fungiert als unumstößlicher Ankerpunkt in der Datenbank-Indizierung, während die spätere Ziffernreihe gerade die Variationen in der Mitte und am Ende des Namens auffängt.

Indem der Anlaut unverändert (aber normalisiert) mitgenommen wird, behält man die notwendige Schärfe beim automatischen Verknüpfen von Nachnamen im Artland-RMS (Artland Record Matching System).

7. Erwartete Ergebnisse und Validierung

Die Implementierung der Artländer Phonetik markiert eine fundamentale Verbesserung bei der automatisierten Verarbeitung und Verknüpfung historischer Regionaldaten. Wo traditionelle algorithmische Standards wie die Kölner Phonetik in einem niedersächsischen Kontext versagen, zeigt die empirische Validierung an dem Datensatz von 175.270 Quellennachweisen eine überzeugende Verschiebung in Effektivität und Präzision.

7.2 Quantitative Ergebnisse: KP versus AP

Der Vergleich zwischen der Nullmessung (Kölner Phonetik) und der Artländer Phonetik zeigt, dass das regionale Algorithmus die Kernprobleme des historischen Namensabgleichs effektiv adressiert:

- **Vollständige Wiedergabe von Varianten:** Während die Kölner Phonetik bei lediglich 8% vollständiger Zuordnungen verharret, gelingt es der Artländer Phonetik, 60% der Varianten innerhalb einer Familiengruppe vollständig und korrekt zu gruppieren. Dies entspricht mehr als einer Versiebenfachung der erfolgreichen, abdeckenden Namensverknüpfungen.
- **Reduktion von Rauschen (False Positives):** Der Prozentsatz verunreinigter Ergebnisse sinkt spektakulär von 36% bei der Kölner Phonetik auf nur noch 8% bei der Artländer Phonetik. Das Phänomen, bei dem nicht verwandte Familien (wie Elting und Oldenhage) zu Unrecht miteinander verknüpft werden, ist damit praktisch ausgebannt.
- **Unvollständige Übersicht:** Die Anzahl unvollständiger Wiedergaben – bei denen relevante Varianten durch zu strenge oder unpassende Regeln übergangen wurden – nimmt von 56% auf 32% ab.

7.3 Qualitative und genealogische Mehrwert

Der Gewinn der Artländer Phonetik lässt sich nicht nur in harten Statistiken erfassen, sondern übersetzt sich direkt in die genealogische Praxis. Durch den geschichteten Aufbau – bei dem nacheinander grafische Normalisierung, regional gebundene Lautgesetze, Stamm-Whitelists und eine geänderte Behandlung des Anlauts angewendet werden – entsteht ein Gleichgewicht zwischen Sensitivität (dem Einfangen echter Varianten) und Spezifität (dem Abwehren unberechtigter Verknüpfungen).

In der täglichen Praxis des Artland Record Matching System (Artland-RMS) bedeutet dies, dass Forscher mit zuverlässigen Vorschlagslisten konfrontiert werden. Wo man früher manuell Tausende von Varianten durchforsten musste oder an den Launen auswärtiger Schreiber und amtlicher Orthographie scheiterte, führt die Artländer Phonetik die zerstückelte historische Realität automatisiert auf den zugrundeliegenden genealogischen Stamm zurück. Die erwarteten Ergebnisse belegen somit, dass ein reginospesifischer, auf Sprachquellen fundierter Algorithmus generischen, modernen Standardlösungen überlegen ist.

Literatur und Kontext

- **Postel, H.J.** (1969). Die Kölner Phonetik. Ein Verfahren zur Identifizierung von Personennamen auf der Grundlage der Gestaltanalyse. In: IBM-Nachrichten, 19. Jahrgang, 1969, S. 925–931.
- **Vehslage, H.** (1908). *Die Mundart des Artlandes auf der Grundlage der Mundart des Kirchspiels Badbergen* (Inaugural-Dissertation zur Erlangung der Doktorwürde der Hohen Philosophischen und Naturwissenschaftlichen Fakultät der Westfälischen Wilhelms-Universität). Buchdruckerei Robert Noske, Borna-Leipzig.