

Hans Geisler

AUFBEREITUNG LEXIKALISCHER INFORMATION UND DEREN UMSETZUNG IN SPRACHKARTEN (AUFGEZEIGT ANHAND ROMANISCHER BEISPIELE)

Wörterbücher und Monographien zu Dialekten oder regionalen Varietäten enthalten in ihren Orts- und Gebietsangaben große Mengen an topographischer Information, die nur durch eine sehr zeitaufwendige Erstellung von Verteilungskarten sichtbar zu machen ist. Für die meisten weiterführenden Untersuchungen muß außerdem die zugehörige lexikalische Information aufgrund der unterschiedlichen Konzeptionen einzelner Lexika und divergierender Verschriftungssysteme erst vereinheitlicht bzw. weiter aufbereitet werden. Es wäre daher wünschenswert, ein Instrumentarium für eine ökonomische Auswertung des betreffenden Materials zur Verfügung zu haben, mit dem gleichzeitig eine optimale kartographische Darstellung möglich wird. Für diese Zwecke wurde ein Programmpaket entwickelt¹, das im wesentlichen folgendes leistet:

- a) einheitliche Datenspeicherung,
- b) flexible Datenselektion,
- c) topographische Darstellung der Daten.

Voraussetzung für eine sinnvolle Verarbeitung von lexikalischer Information mit Hilfe eines Computers ist, daß sowohl am Bildschirm als auch auf dem Drucker die Zeichen des verwendeten Transkriptionssystems zur Verfügung stehen. Im konkreten Fall wurde deshalb mit einem Fonteditor auf der Basis einer Antiquaschrift eine für den romanischen Bereich ausreichende Untermenge des IPA-Zeichenvorrats erstellt und mit einer darauf abgestimmten Tastaturbelegung gekoppelt.

Das in IPA-Transkription umgesetzte Material kann somit ohne weitere Umcodierung in einer speziell konzipierten Datenbank editiert und verwaltet werden. Damit entfällt die in Verbindung mit normalen Datenbanken notwendige und fehlerträchtige Eingabe von komplexen Ersatzkodierungen. Die in der Datenbank gespeicherten Daten können mittels einer einfachen "Abfragesprache" nach bestimmten linguistischen Kriterien selektiert werden. Mit den Ausgangsdaten oder den selektierten Daten lassen sich Ortspositionen, Hintergrundkarten und Legendentexte verbinden, was eine Ausgabe in Kartenform erlaubt, sei es als Rohdatenkarte, Typenkarte oder Isoglossenkarte. Die Übernahme der Daten in Textverarbeitungssysteme (z.B. für eine formatierte Ausgabe in Listenform) ist ebenfalls möglich.

Die Möglichkeiten des Programms sollen exemplarisch anhand einiger Beispiele aus dem romanischen Bereich aufgezeigt werden. Die lexikalischen Ausgangsdaten sind dem *Französischen Etymologischen Wörterbuch (FEW)* von W. von Wartburg entnommen, da es bereits mehr als 1000 einzelne

Dialektwörterbücher aus dem galloromanischen Bereich verarbeitet hat und damit hervorragend als Grundlage für topographische Untersuchungen geeignet ist. Das *FEW* ist konzipiert als "darstellung des galloromanischen sprachschatzes" (s. Untertitel) und will die Geschichte aller galloromanischen Wortfamilien mit allen Bedeutungen und allen Ableitungen seit dem Beginn bis ins Neufranzösische, eingeschlossen alle Dialekte Frankreichs, der Wallonie, des Frankoprovenzalischen in Frankreich, der Schweiz und Italien und des Okzitanischen nachvollziehen². Das gesammelte Material aus literarischen Quellen, Wortsammlungen, Vokabularen, Dialektatlanten etc. ergibt einen Lexikonumfang von einigen Millionen Wörtern und Wortformen in rund 20 000 Artikeln (über 20 Bände). Diese Materialfülle ist bisher leider nur durch Teilindices erschlossen und deshalb nur schwer zugänglich.

Besonders nachteilig wirkt sich aufgrund der Fülle des Materials aus, daß bei bestimmten Lemmata die Verteilung der einzelnen Formen und Bedeutungen nur mehr sehr schwer zu überblicken ist. W. von Wartburg hat es daher selbst als dringendes Bedürfnis empfunden, ausgewählte Artikel durch Kartenbeigaben zu illustrieren, sei es um die Ausdehnung der verschiedenen Benennungen eines Begriffs sichtbar zu machen oder die Verteilung phonetischer Typen aufzuzeigen. Vor allem die phonetische Information ist aufgrund der Konzeption auf verschiedene Stellen der Artikel verteilt, und ihre Auswertung erfordert eine Neuorganisation des Materials. Der Plan, zu einzelnen Artikeln Karten zu erstellen, der zur damaligen Zeit aus Kostengründen aufgegeben werden mußte, kann nun nachträglich mit dem vorgestellten Programm realisiert werden.

Zunächst muß mit Hilfe des Zeichenprogramms eine Hintergrundkarte von Frankreich mit Départementgrenzen gezeichnet werden, in die die Ortskoordinaten der Aufnahmepunkte einzutragen sind. Im konkreten Fall sind dies alle im *FEW* verwendeten Ortspunkte in der Zitierreihenfolge des Wörterbuchs (s. *FEW Beiheft* 1950: 63ff. und *FEW Supplement* 1957:11ff.). Das *FEW* übernimmt zusätzlich alle Aufnahmepunkte des *ALF* und einige piemontesische Punkte des *AIS*, die ebenfalls in die Grundkarte aufgenommen werden. Es ergibt sich damit für den galloromanischen Bereich eine flächendeckendes Aufnahmenetz von ungefähr 1500 Punkten (teilweise mit Punktabständen unter 30 km)³. Eine Einarbeitung neueren Materials sowie von Regionalatlanten (z.B. der *ALF par régions*) wäre ergänzend möglich. Bei einer derartigen Aufnahmedichte ist es nicht sinnvoll, Rohdaten direkt auszugeben. Es müssen vielmehr Typisierungen vorgenommen werden, die in Symbole umgesetzt werden können. Ein entsprechender Symbolfont ist mit Hilfe des Fonteditors zu erstellen.

Um eine Weiterverarbeitung des Materials vor allem aus phonetischer Sicht zu ermöglichen, muß bei der Eingabe eine Vereinheitlichung der Transkription auf das IPA-System erfolgen, da im *FEW* sowohl verschiedene Orthographiesysteme als auch ein an Böhmer-Ascoli angelehntes Transkriptionssystem Verwendung findet. Falls noch weitere Quellen einge-

geben werden (z.B. der *ALF* oder die *ALF par régions*, mit jeweils voneinander abweichenden Transkriptionssystemen), müssen diese ebenfalls auf das IPA-System umgeschrieben werden.

Zwischen den verschiedenen Systemen sind demnach zuerst Äquivalenzbeziehungen zu definieren, die eine möglichst eindeutige Überführung in das IPA-System erlauben. Dies kann in Einzelfällen problematisch sein, da für die dialektalen Orthographien nicht immer eindeutige Lautwerte bestimmbar sind⁴.

Ein weiteres Problem besteht darin, daß im Gegensatz zu Sprachatlanten die Daten des *FEW* nicht alle gleichgeordnet sind. Dialektformen können in hierarchischer Beziehung zueinander stehen (z.B. Ort – Gegend – Dialekt – Sprache) oder verschiedenen Zeitstufen angehören (altfranzösisch – mittelfranzösisch – neufranzösisch). Derartige Inklusions- oder Transformationsbeziehungen können in bedingtem Umfang durch typographische Kennzeichnungen sichtbar gemacht werden⁵.

Die Auswertung lexikalischer Einträge des *FEW* und *LEI* sollen folgende Beispiele illustrieren:

In Beispiel 1 wird unter onomasiologischem Gesichtspunkt Lexikonmaterial aus *FEW* und *LEI* neu gruppiert, um die Bezeichnungen für die 'Himbeere' in ihrer Verteilung in Frankreich und Norditalien sichtbar zu machen. Als Ordnungsprinzip dient im *FEW* die etymologische Ausgangsform einer Wortfamilie (d.h. um ein Wort überhaupt zu finden, muß das Etymon bekannt sein.). Die weitere Untergliederung erfolgt nach den verschiedenen Bedeutungen, welche die Wortfamilie angenommen hat (zum formalen Gesamtaufbau s. Anhang 1). Unmöglich war die Ordnung nach Etyma bei nicht etymologisierbaren Wortformen, weshalb diese in drei separaten Bänden (Bd. XXI, XXII, XXIII "Materialien unbekannten oder unsicheren Ursprungs") nach onomasiologischen Prinzipien angeordnet sind. Zugrunde liegt hierbei das Begriffssystem von Hallig – Wartburg. Aufgrund dieses Anordnungsprinzips sind die Bezeichnungen für die 'Himbeere' auf zwei Artikel verteilt. Der französisch weit verbreitete Typ 'framboise' wird auf ein germanisches Etymon **brambasi* zurückgeführt und erscheint in *FEW* XV,1 (Germanische Elemente). Die restlichen Bezeichnungen wurden von W. von Wartburg als nicht eindeutig etymologisierbar erachtet und deshalb in *FEW* XXI "Materialien unbekannten oder unsicheren Ursprungs" unter dem Begriff 'framboise' aufgelistet. Um einen Eindruck vom Ausgangsmaterial des *FEW* zu vermitteln, wird in Anhang 2 ein Auszug der betreffenden Artikel wiedergegeben. Anhang 3 zeigt die normierten Daten wie sie in der Datenbank gespeichert sind und als Grundlage für die Datenabfrage und Symbolzuordnung dienen. Die daraus resultierende Karte zeigt gut die Verteilung der einzelnen Typen (s. Karte 1). Der vorromanische Typ **amp* setzt sich auf italienischem Gebiet fort, wie die aufgrund der Daten des *LEI* zu **amp* erstellte Verteilungskarte verdeutlicht (s. Karte 2; weitere Bezeichnungen sind dem *LEI* nicht zu entnehmen, da erst Bd. 1 und 2 erschienen sind).

Beispiel 2 zeigt unter semasiologischem Gesichtspunkt die Bedeutungsentfaltung der Fortsetzer von *Maurus* in der Galloromania (s. Karte 3). Dieser Kartentyp ist aufgrund der Konzeption des *FEW* relativ leicht zu erstellen, da das Material bereits entsprechend geordnet ist. Die Karte dient also im wesentlichen dazu, die Verteilung der einzelnen Bedeutungen genau sichtbar zu machen.

Beispiel 3 wertet die im *FEW* enthaltene phonetische Information aus, die nur durch eine vollständige Neuordnung der in einem Artikel enthaltenen Wortformen sichtbar gemacht werden kann. Die Karten zu *salice* (s. Karten 4 und 5) zeigen, in welchem Ausmaß die im *FEW* und weiteren Quellen enthaltene Information bereits bestehende Atlanten wie den *ALF* ergänzen und korrigieren kann, und damit phonetische Entwicklungen (hier die Ausdehnung der Synkope) präziser beschreibbar macht. Beispiel 4 zeigt die Verteilung bestimmter Synkopeformen für *basilica* (s. Karte 6). Für diese Karte existiert keine Entsprechung im *ALF*.

Diese wenigen Beispiele lassen erkennen, wie die immense Materialfülle eines *FEW*, *LEI* oder ähnlich konzipierter Lexika mit Hilfe des vorgestellten Programms für Spezialuntersuchungen in verschiedenen Bereichen nutzbar gemacht werden kann.

FUSSNOTEN:

1. Die Programme wurden von D. Strehle (München) in Turbo Pascal und Assembler geschrieben und können auf einem PC mit Graphikkarte (EGA, VGA) in Verbindung mit einem Lasedrucker verwendet werden. Ausführlicher zum Programm s. Geisler 1990.
2. Entsprechende Zielsetzungen für den italienischen Sprachraum setzt sich das *Lessico etimologico italiano* (*LEI*); s. Pfister 1973:245 "Das grundlegende Ziel des italienischen etymologischen Wörterbuches besteht darin, den gesamten erreichbaren Wortschatz der italienischen Schriftsprache und der Dialekte Italiens, Korsikas und der Südschweiz aus gedruckten Quellen systematisch aufzunehmen und wissenschaftlich zu interpretieren".
3. Ungefähr 150 Punkte des *FEW* stimmen mit *ALF*-Punkten überein, werden aber extra geführt, da das *FEW* zusätzliche Quellen verwendet.
4. Aus typographischen Gründen ist es nicht zweckmäßig, die sehr enge "impressionistische" Transkription des *ALF* und *AIS* beizubehalten. Als Vorbild zur Vereinfachung dienen dem *FEW* wie auch dem *LEI* die von Jaberg/Jud (1928:36) gegebenen Empfehlungen.
5. In der Abfrage können hier Gruppen gebildet werden, welche diese Beziehungen zum Ausdruck bringen. So kann z.B. vereinbart werden, daß bei Zugehörigkeit eines Ortes zu einem Dialekt, bei fehlenden Ortsdaten, die entsprechenden Dialektdaten Geltung haben.

6. Ein Vergleich mit *ALF* 609 läßt erkennen, daß die *ALF*-Karte nur einen Teil der Bezeichnungen erfaßt (wohl weil oft der französische Name für die Kulturhimbeere genannt wurde).

BIBLIOGRAPHIE:

- AIS = Jaberg, K., Jud, J., Sprach- und Sachatlas Italiens und der Südschweiz, 8 Bde., Zofingen 1928-1940.
- ALF = Gilliéron, J., Edmont, E., Atlas linguistique de la France, 10 Bde., Paris 1902-04.
- Baldinger, K., 1974, "Le *FEW* de Walther von Wartburg", in: Introduction aux dictionnaires les plus importants pour l'histoire du français, Paris, 11-47.
- FEW = Wartburg, W.v., Französisches Etymologisches Wörterbuch. Eine Darstellung des galloromanischen Sprachschatzes, Bonn - Leipzig - Tübingen - Kassel, 1922-.
- FEW Beiheft = Wartburg, W.v., Französisches Etymologisches Wörterbuch. Eine Darstellung des galloromanischen Sprachschatzes, Beiheft (Ortsnamenregister - Literaturverzeichnis - Übersichtskarte), Tübingen 1950 [1929].
- FEW Supplement = Wartburg, W.v., Französisches Etymologisches Wörterbuch. Eine Darstellung des galloromanischen Sprachschatzes, Supplement (zur 2. Auflage des Bibliographischen Beiheftes), Basel 1957.
- Geisler, H., 1990, "Erstellung und Auswertung von Dialektkarten mit Personal Computern", Akten der Sektion Computerlinguistik des XXI. Romanistentages, RWTH Aachen 25.-27.9.1989 (ed. J. Rolshoven), Tübingen, Niemeyer, ersch.
- Jaberg, K., Jud, J., 1927, "Transkriptionsverfahren, Aussprache- und Gehörsschwankungen", ZRPh 47, 171-218.
- Jaberg, K., Jud, J., 1928, Der Sprachatlas als Forschungsinstrument. Kritische Grundlegung und Einführung in den Sprach- und Sachatlas Italiens und der Südschweiz, Halle.
- LEI = Pfister, M., Lessico etimologico italiano, Wiesbaden, 1979-.
- Pfister, M., 1973, "Das Projekt eines italienischen etymologischen Wörterbuches (IEW)", ZRPh 89, 244-272.

ANHANG

Anhang 1: Formales Grundgerüst eines *FEW*-Artikels (nach Baldinger 1974:27ff.)

TITELWORT	(lat., germ. etc. Etymon)
I. Erbwörtliche Formen	
Bedeutung 1, 1a, 1aa, 1aaa', 1aaa'a", α"; 1b, 1ba ...	
Basalwort	(afr...wall...gask., s. Zitierordnung im Beiheft)
Redewendungen	
Übertragene Bedeutungen	
Ableitungen	(nominal, verbal, deverbale)
Zusammensetzungen	
Bedeutung 1a, 1aa, 1aaa', 1aaa'a", α"; 1b, 1ba ...	
Bedeutung 2, 2a, ...	
...	
II. Gelehrte Formen	(aus dem Lat., aus anderen Sprachen)
Kommentar	(mit Erklärungen zur Struktur, Genealogie etc.)
Fußnoten	

Anhang 2: Originaldaten des *FEW* zu 'Himbeere'. (Der Auszug gibt typographisch das Original wieder; Transkription wird durch Weitstellung angezeigt; Dreiecke weisen auf Auslassungen)

<FEW XV,1,239>

*brambasi (anfrk.) brombeere

1.1. **Himbeere.** - Fr. *framboise* f. "fruit du framboisier" (seit BenSmaure), lütt. *framwèse*, Bastogne *f r ā b ā š y*, NivB. *f l ā b q š*, MarcheEc. *flambéje*, Ath *f l ā b q s*, Mons *flambèse*, rouchi *flambesse*,
... ▲ ...

<FEW XXI,1,93>

framboisier; framboise

1. Mfr. *ambre* "framboise" (Rethel 1408), Neufch. id., ardch. Guign. Landres id., mouz. argonn. *ambe*, Marne *ā b* (p 155), Meuse *ā b* (p 156), argonn.

Vouth. Chatt. Gruey, Brotte, bress. Plancher, Belf. *ambre*, Châten. *ômbre*, Montbél. *ambre*, Dampr. *5 b r*, Ajoie, FrMont. *â b r*, Barc. *ambra* "mûre sauvage",

... ▲ ...

3. a. Malm. *a m ô n f*. "framboise", lütt. *d m ô n* "id.; mûre", Seraing *o m o n*, Marche *a m 5 t*, "framboise", nam. ardw. *a m â d*, War. *a m 5 n*. - Ablt. Faymonv. *âmounîre*, Lütt. *âmoni* "framboisier",

.. ▲ ...

... Es liegt wohl ein vorromanisches wort vor, dessen Stamm in drei lautlichen varianten vorliegt: *amb-* (1), *amp-* (2), *amm-* (3). 1, das mit einem -r-suffix versehen ist,

... ▲ ...

1. Aude *z é r s i m*. (p 793), Ariège *z é r s e* (p 792), *z í r s e* (p 791), Prades *z é r s* (pl. *z é r s i*), Ascou *a z é r s* (pl. *a z é r s i*), Suc, Sentenac *jèr s e s* pl. Bts 1933, 58. - Ablt. Prades ..

... ▲ ...

Zu 1 vgl. PyrOr. *y é r s u*, *z é r s u* ALF 609, gers, ursprünglich plural, kat. *gerd*; zu 2 Roncal, arag. *chordón*, Ribagorza *jordón*⁵). Daraus lassen sich als grundformen iber. **jirde*, **jurdone* erschliessen;

...▲...

1. Bearn. *mastà m*. "lieu planté de framboises".

2. Bearn. *mastajoû*, Lescun *m a s t a y û*. - Ablt. Bearn. *mastajoè* "lieu planté de framboises",

... ▲ ...

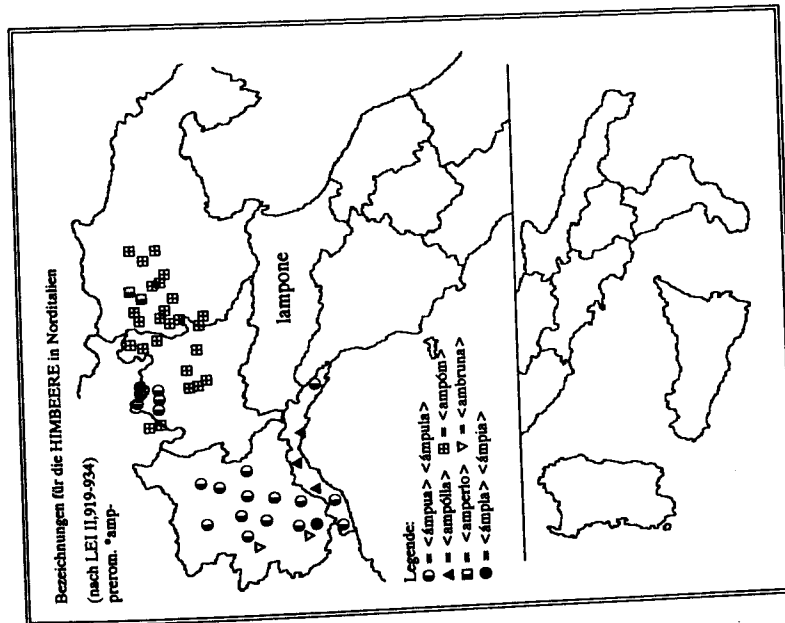
Auszugehen ist von iber. **mas-ta* "himbeere", daher rom. **mastaris* > bearn. *mastà*

... ▲ ...

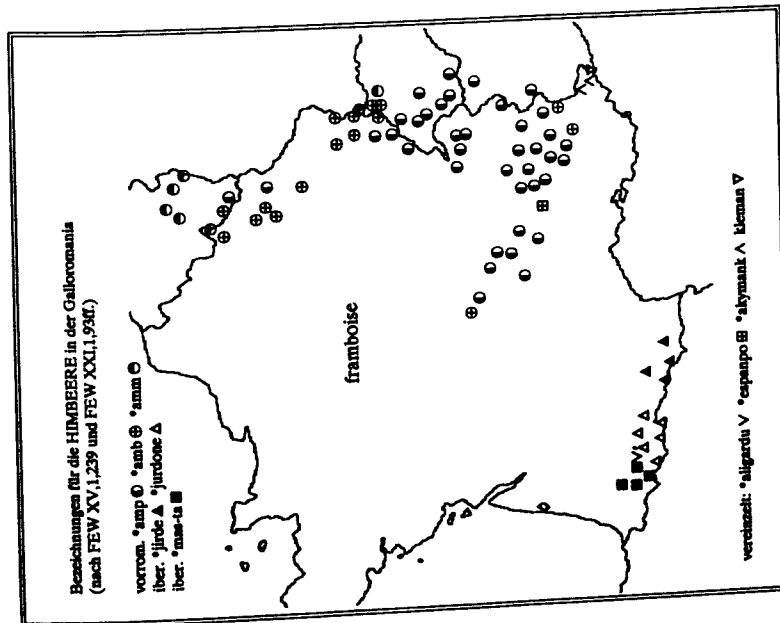
Anhang 3: FEW-Daten in der Datenbank (dazwischenliegende Orte ohne Eintrag sind ausgelassen)

Ort	FEW-Form	IPA-Form	...	Typ
fr.	<i>framboise</i>	frābwaz	...	*brambasi
Malm.	<i>a m o n</i>	amo:n	...	*amm
Seraing	<i>o m o n</i>	omon	...	*amm
Marche	<i>a m ʒ t</i>	amʒt	...	*amm
Bastogne	<i>f r ā b ā ʒ y</i>	frābafj	...	*brambasi
Neufch.	<i>ambe</i>	ābə	...	*amb+r _{suf}
NivB.	<i>f l ā b o ʒ</i>	flāboʒ	...	*brambasi
Ath	<i>f l ā b o s</i>	flābos	...	*brambasi
ardw.	<i>a m ā d</i>	amād	...	*amb+r _{suf}
mouz.	<i>ambe</i>	ābə	...	*amb+r _{suf}
argonn.	<i>ambe</i>	ābə	...	*amb+r _{suf}
Marne 155	<i>ā b</i>	āb	...	*amb+r _{suf}
Ajoie	<i>ā b r</i>	ābr	...	*amb+r _{suf}
Aude 793	<i>ʒ ɛ r s i</i>	ʒersɪ	...	*jirde
Ariège 792	<i>ʒ ɛ r s e</i>	ʒerse	...	*jirde
Prades	<i>ʒ ɛ r s</i>	ʒers	...	*jirde
Ascou	<i>a ʒ ɛ r s</i>	aʒers	...	*jirde
bearn.	<i>mastà</i>	mas'ta	...	*mas-ta

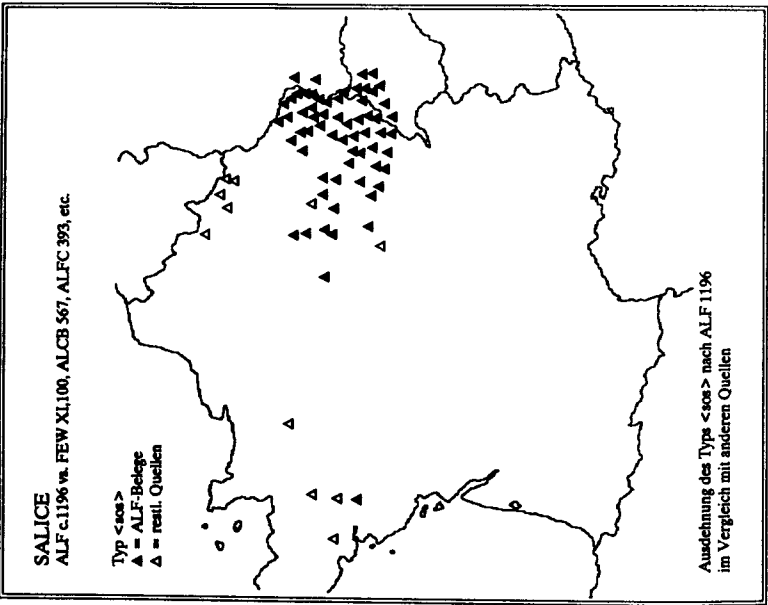
Karte 2



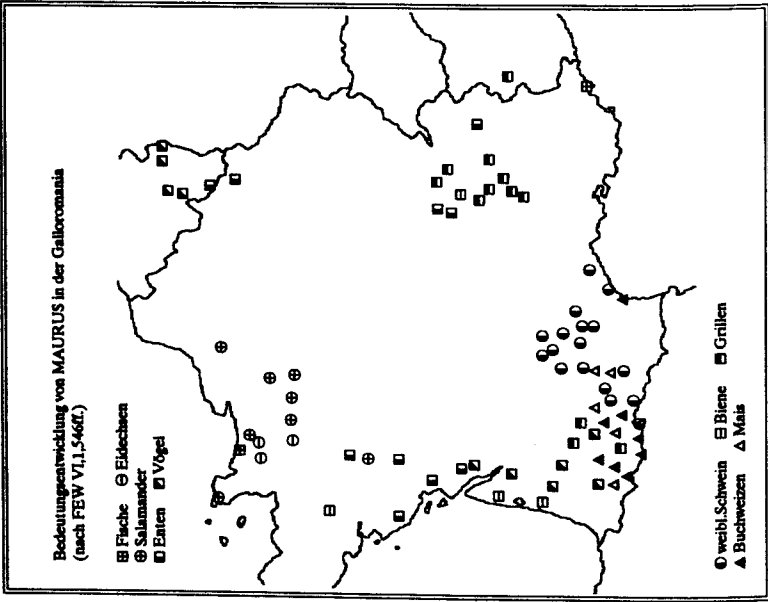
Karte 1



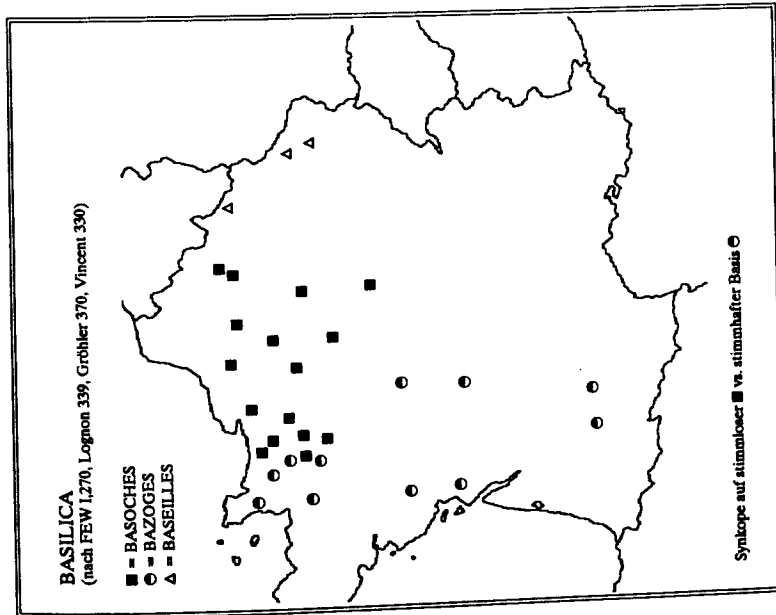
Karte 4



Karte 3



Karte 6



Karte 5

