

*TE'UDA*  
THE CHAIM ROSENBERG SCHOOL  
OF JEWISH STUDIES RESEARCH SERIES

XVIII

SPEAKING HEBREW  
Studies in the Spoken Language  
and in Linguistic Variation in Israel

*Editor*  
SHLOMO IZRE'EL

*With the Assistance of*  
MARGALIT MENDELSON

TEL-AVIV UNIVERSITY • 2002

*Editorial Board of the Series*

Yair Hoffman, Chairman	Yoram Yacobson
Nurit Govrin	Moshe Florentin
Eliezer (Ed) Greenstein	Tova Rosen
Itamar Singer	Elhanan Reiner

- Volume 1:* Cairo Geniza Studies, 1980  
*Volume 2:* Studies in Bible, 1982  
*Volume 3:* Studies in Talmudic Literature, Post-Biblical Hebrew and Biblical Exegesis, 1983  
*Volume 4:* Studies in Judaica, 1986  
*Volume 5:* Studies in Hebrew Literature, 1986  
*Volume 6:* Studies in Hebrew and Arabic, 1988  
*Volume 7:* Studies in Judaica, 1991  
*Volume 8:* Studies in the Works of Abraham Ibn Ezra, 1992  
*Volume 9:* Studies in Hebrew Language, 1995  
*Volume 10:* Studies in Judaica, 1996  
*Volume 11:* Studies in the Aggadic Midrashim 1996  
*Volume 12:* The Jewish Diaspora in the Hellenistic and Roman Periods, 1996  
*Volume 13:* Marriage and the Family in Halakha and Jewish Thought, 1997  
*Volume 14:* Encounters in Medieval Judaeo-Arabic Culture, 1998  
*Volume 15:* A Century of Geniza Research, 1999  
*Volume 16–17:* Studies in Judaica, 2001

©

By Tel Aviv University  
Printed in Israel, 2002  
by Graphit Ltd., Jerusalem

TEL AVIV UNIVERSITY  
THE LESTER AND SALLY ENTIN FACULTY OF HUMANITIES  
THE CHAIM ROSENBERG SCHOOL OF JEWISH STUDIES  
RESEARCH SERIES

PUBLISHED WITH THE ASSISTANCE OF THE MOSHE DORF FOUNDATION



**Moshe Dorf**

(June 2, 1907 — October 5, 1987)

Moshe Dorf was born in Poland and emigrated to Belgium at an early age. Together with his brother, he founded a diamond processing enterprise. The personality of Moshe Dorf reflects the embodiment of initiative and action which received their concrete expression in the industrial enterprise that he established and continued to cultivate throughout his life. Concurrently, he possessed an intense spiritual wealth as well as an unquenchable thirst for knowledge and wisdom from the rich sources of the Jewish heritage. He was an active participant in various study groups and conferences devoted to Jewish Studies, especially in the field of biblical research, and an avid member of the Israeli Society for Biblical Research as well as the World Jewish Society for the Study of Bible.

Moshe Dorf will be remembered by friends and relations for his congeniality and amiability, his unstinted devotion to his family, and his munificent philanthropy to numerous cultural and charitable institutions in the State of Israel.





## TABLE OF CONTENTS \*

FORWARD	by Yair Hoffman, Editor, the <i>Te'uda</i> series	87
PREFACE	by Shlomo Izre'el, Editor of <i>Te'uda</i> 18	17

### Corpus Linguistics and Computational Linguistics

John Sinclair		
CORPUS LINGUISTICS: THE STATE OF THE ART		3
John Sinclair		
LEXICAL GRAMMAR: A NEW LOOK AT LANGUAGE		15
Shuly Wintner		
HEBREW COMPUTATIONAL LINGUISTICS: PAST AND FUTURE		35

### Language and Society in Israel

Eliezer Ben-Rafael		
MULTICULTURALISM AND MULTILINGUALISM IN ISRAEL		67
Muhammad Amara		
HEBREW AMONG THE ARABS IN ISRAEL: SOCIOLINGUISTIC ASPECTS		85
Otto Jastrow		
<i>THE CORPUS OF SPOKEN PALESTINIAN ARABIC (COSPA)</i>		107

\* The articles marked with an asterisk are products of the research workshop of *The Corpus of Spoken Israeli Hebrew (CoSIH)* (Atlanta 2000).

Elana Shohamy and Bernard Spolsky	
FROM MONOLINGUAL TO MULTILINGUAL? EDUCATIONAL LANGUAGE POLICY IN ISRAEL	115

### **Linguistic Variation**

Yaakov Bentolila	
LINGUISTIC VARIATION ACROSS GENERATIONS IN ISRAEL	131
Ora (Rodrigue) Schwarzwald	
LANGUAGE VARIETIES IN CONTEMPORARY HEBREW	141
Zohar Livnat	
ON LANGUAGE, LAW, AND SOCIAL JUSTICE	177

### **Spoken Hebrew in Israel and It's Study**

Moshe Bar-Asher	
MODERN HEBREW AND ITS CLASSICAL BACKGROUND	203
Shlomo Izre'el	
THE EMERGENCE OF SPOKEN ISRAELI HEBREW	217
Shmuel Bolozky	
PHONOLOGICAL AND MORPHOLOGICAL VARIATION IN SPOKEN HEBREW	239
Geoffrey Khan	
THE STUDY OF MODERN HEBREW SYNTAX	279
Yael Reshef	
THE SOCIOLINGUISTIC PHENOMENON OF V FORM IN HEBREW IN THE BRITISH MANDATE PERIOD	299

Ron Kuzar	
THE SIMPLE IMPERSONAL CONSTRUCTION IN TEXTS REPRESENTED AS COLLOQUIAL HEBREW	329
Esther Borochofsky-Bar Aba	
BETWEEN SPOKEN AND WRITTEN LANGUAGE: EXAMINING PARALLEL SPOKEN AND WRITTEN TEXT	353
Yitzhak Shlesinger	
POLARITY IN LANGUAGE LEVELS IN LITERARY TEXTS	375
Tamar Sovran	
SPOKEN AND POETIC LANGUAGE IN ISRAELI MODERN POETRY	395
Il-Il Yatziv	
FROM TRANSCRIPTION OF SPOKEN TEXT TO ITS REPRESENTATION ON A GRID SET	421

**Toward the Compilation of *The Corpus of Spoken Israeli Hebrew*  
(CoSIH)**

Giora Rahav	
POPULATION SAMPLING FOR THE ESTABLISHMENT OF A REPRESENTATIVE CORPUS	439
Benjamin Hary and Shlomo Izre'el	
THE PREPARATORY MODEL OF <i>THE CORPUS OF SPOKEN ISRAELI HEBREW (COSIH)</i>	447
Regina E. Werum	
METHODOLOGICAL REMARKS ON CREATING <i>THE CORPUS OF SPOKEN ISRAELI HEBREW (CoSIH)</i>	459
ENGLISH ABSTRACTS	7*



## FOREWORD

Of the 17 volumes of *Te'uda* that have preceded this one, 8 are interdisciplinary in nature, reflecting a wide range of research topics in Jewish Studies. The other volumes each focused on a major topic: the Cairo genizah (1; 15); the Bible (2); the Talmud, Biblical commentary, and Jewish Thought (3; 13); Hebrew literature (5); Hebrew and Arabic (6; 9); and Jewish-Arabic cultural encounters in the Middle Ages (14). The eighteenth volume of *Te'uda*, which we are happy to present to our readers, belongs to this second group, but it also introduces an innovative note: This is the first volume that does not focus on written texts and is entirely devoted to a research topic that is still in its infancy and deals with the present. This represents a statement, both academic and cultural: Even in the field of Jewish Studies it is important and possible to study not only written texts, but also the present and the evolving, which have not yet been immortalized in writing; the Hebrew language that is taking form. What belongs more to this category than spoken language? It is an important topic of research, which goes beyond the narrow realm of linguistic study to a study of Israeli society in its entirety. The living, spoken Hebrew language is one of the most striking achievements, some maintain the most important achievement, of the Israeli experience, because it is at once a formative factor and a mirror of the society in which we live. As such, the Hebrew language is deserving of a research tool, research methods, and scholars who will develop them. I hope that this volume will serve to advance these goals.

Yair Hoffman, Series Editor

## PREFACE

The subject of this volume being offered to scholars of the Hebrew language and its devotees is an exploration of spoken Hebrew. Its nucleus is a meeting of scholars that occurred in February 2000 at Emory University in Atlanta (Georgia, USA). Papers were delivered by team members of *The Corpus of Spoken Israeli Hebrew (CoSIH)* and other scholars invited for this purpose, and questions were discussed pertaining to the compilation and design of the corpus. At the initiative of Yair Hoffman, editor of the *Te'uda* series, other scholars engaged in the research of the Hebrew language were asked to contribute from their research to this volume. These articles, to a large extent, supplement the writings of the CoSIH workshop seminar, because they show — from various and diverse aspects — the pressing need for compiling a corpus of spoken Israeli Hebrew.

The study of spoken Hebrew is still in swaddling clothes, perhaps even in gestation. Without a corpus of data, there cannot be any research. Random data retrieval enables observation and insights. Authentic research that stems from research hypotheses and includes an examination of the data, followed either by corroboration or by refutation of the hypotheses (not only theory for its own sake), is possible only in a limited way when the research has no proper data base. Therefore, this volume should be properly valued for what it is: a proposal for directions of research into spoken Hebrew.

The research workshop in Atlanta consisted of several sections, and this publication essentially preserves that organization. The first section deals with the methodology of language research according to corpora. The section opens with two articles by John Sinclair, a professor at the University of Birmingham (England) and the founder and director

(together with Elana Tognini-Bonelli) of the Tuscan Word Center in Italy. John Sinclair is one of the founding fathers of the linguistic method that sees the language corpus as a fitting basis for a new view of language. The first article presents the reader with the state of the art in corpus linguistics; in his second article, there is a survey of the achievements of research in corpus linguistics in the innovative field of “lexical grammar”, a field of research in which Sinclair’s own contribution is definitive. The tremendous momentum accorded to language research supported by language corpora would not have been possible without the major developments in the computer sciences and computerization technology of the past decades. In general, no comprehensive linguistic research today is possible without the use of computers. Shuly Wintner, a computational linguist, was asked to survey the state of the art in this field in the study of Hebrew, and his article concludes the first section.

The second section consists of four articles and presents a profile of the relationship between language and society in Israel. Eliezer Ben-Rafael, President of the International Institute of Sociology and President of the Israeli Association for the Study of Language and Society, reviews the cultural and linguistic variety in our country. Muhammad Amara, a researcher of Arabs in Israel and their languages, turns his scrutiny to the Hebrew spoken by Arabs in Israel and asks about the attitude of Arabs toward Hebrew. Otto Jastrow, an expert in Arabic dialectology, reviews the status of the compilation of *The Corpus of Spoken Palestinian Arabic (CoSPA)*. Elana Shohamy and Bernard Spolsky, whose expertise is in the linguistic policy, in language education and issues of language and society, discuss the policy of language education in Israel.

The third section contains three articles that deal with linguistic variation. Language variation may be the result of demographic variance or the result of language use in different discourse contexts. Yaakov Bentolila, a distinguished scholar of spoken Hebrew and Jewish languages, and one of the first to recognize the need to examine spoken Hebrew as a subject of research in its own right, shows that age, a demographic variable, is a dominant variable in language use. Ora Schwarzwald, one

of the leaders in the study of Israeli Hebrew, proposes a classification of language varieties according to either demographic or contextual criteria. This section concludes with an article by Zohar Livnat that shows the need to study language varieties and their relationships as she observes the language of law and its users.

The fourth section is devoted to a study of language itself. The first article is by Moshe Bar-Asher, President of the Academy of Hebrew Language and an important scholar of Hebrew and Jewish languages, who shows how Modern Hebrew continues the tradition of past generations in its creative aspect. The creative aspect is demonstrated, among other ways, by the uses of spoken Hebrew — uses that would no longer be recognized as novel and, therefore, constitute evidence that they are essential to the full and complete vitality of our language. In the second article of the section, I propose a new approach to observing the formative processes of spoken Israeli Hebrew in the early twentieth century. Shmuel Bolozky, an eminent scholar of contemporary Hebrew, conducts a comprehensive, detailed survey of phonological and morphological variation in spoken Hebrew, a survey that draws largely upon his own long-term and productive occupation with the subjects of his research. Geoffery Khan, a specialist in Semitic linguistics and the syntax of Semitic languages, reviews the state of research into Modern Hebrew syntax. The next five articles test observation methods in the relationship between the spoken language and the written language. Yael Reshef examines the formal terms of address (V form) in correspondence during the British mandate, which she found in the archives of the City of Tel Aviv, and tries to evaluate whether this stemmed from a spoken level of Hebrew of some kind. Yitzhak Shlezinger, who has dealt at length with language varieties in the press, takes a look, in his article, at the polarity between the language used in narratives and dialogues in Hebrew literature, a form of language that tries to represent spoken language. Ron Kuzar investigates the simple impersonal construction in language represented in literature as spoken language. Tamar Sovran presents the entry of spoken Hebrew into the Hebrew poetry of recent years and examines the sources of this change. Esther



Borochovski-Bar Abba evaluates the differences between parallel texts that appear in speaking and writing and draws preliminary conclusions about the unique characteristics of the spoken language. When we are in possession of a large corpus of spoken Hebrew, says Kuzar, it will be possible to examine the validity of ideas proposed here on the basis of a solid methodology, and this is true of any investigation of spoken language using the data available today. The last article in this section is by Il-Il Yatziv, a student of Claire Blanche-Benveniste, an outstanding linguist working on spoken French, who developed a fascinating method for investigating the syntax of a spoken language. Yatziv applied this method to Hebrew and presents its principles here.

The last section presents reviews of the methodology to be employed in the compilation of *The Corpus of Spoken Israeli Hebrew (CoSIH)*. Giora Rahav, who is responsible for the statistical and social aspects of the *CoSIH* framework, writes about population sampling and establishing a representative corpus of language. Benjamin Hary and I, the initiators of *CoSIH*, who together headed the planning team, present the corpus design, as presented before the *CoSIH* team in February 2000. This design is the design of a model before application. A methodological perusal of this design by Regina Werum, from a sociological viewpoint, concluded the *CoSIH* workshop and concludes this volume.

My warmest thanks to my colleague Benjamin Hary, who must be credited with the success of the Atlanta conference and the launching of the long-term *CoSIH* project. My thanks to members of the *CoSIH* team for their participation and support. Thanks to Giora Rahav for his help in editing the articles that deal with social questions. Thanks to Yair Hoffman, Director of the Rosenberg School of Jewish Studies and to members of the school's publication committee, who recognized the project's importance and supported the publication of the proceedings of the workshop in this expanded and admirable format. In conclusion, kudos to Margalit Mendelson, who worked so devotedly with me in editing this volume.

...dispersed with the text if and when required in the course of an application. ... is irredeemably corrupted

Another practice that should be avoided, given what is entailed in the size issue, is the manual analysis of corpora. Although there are drawbacks to automatic analysis, there are also important gains, and since the corpora of the future will be too large to handle manually, the effort of analysis should be focused more and more on the automatic process.

In the world of information science outside linguistics, it is necessary to discover something of the meaning of texts in order to arrange for their retrieval by automatic search engines. The prevailing view in this scientific area is that language is not a sufficient self-organising medium, and some other model of information has to be applied. There are many models and hundreds of research groups trying to improve access to language text because most of the internet is presented in written language. So far, linguists have failed to show that language is sufficiently organised to be approached directly. This point links back to the reluctance to explore automatic analysis, and to the issue of size.

The ultimate question is whether a computer could eventually be programmed to comprehend natural language text in a similar way to a human user of the language. So far no genuinely intractable feature of language has been adduced, though everyone agrees that it is a formidable task.

## **LEXICAL GRAMMAR: A NEW LOOK AT LANGUAGE**

© John Sinclair 2002

...tics traditionally divides the meaning-making patterns of language into two types. One, usually called *grammar* or *syntax*, concentrates on the abstract arrangements of words into phrases, phrases into sentences, etc. The particular quality of individual words is played down and

ignored as much as possible. Words are divided into those common ones that regularly indicate grammatical constructions, and the rest, whose meaning is handled in dictionaries, where the grammatical disposition of the words is an occasional and minor matter. This type of meaning making is usually called *semantics*.

The early findings from the analysis of large language corpora are that this very basic distinction may not be necessary — in fact it may act as a distorting lens for researchers. Dividing the types of pattern weakens both of them; it becomes impossible in practical terms for a grammar to state comprehensively the classes that operate in structures, and it becomes impossible to describe comprehensively the meanings of words since so many meanings involve word combinations which in turn involve grammatical structures.

It is accepted that the meanings associated with grammar are not confined within it, but can also be created by semantic processes, and by the combination of semantics and grammar. Polarity, modality, tense etc. are examples. Also the description of pragmatic and similar meanings often requires that words in combination are assigned meanings that they do not normally have when examined individually.

Corpora are oriented to the study of words and word combinations rather than abstractions, and so it is natural for lexically oriented descriptions to emerge at the early stages of research with this novel resource. As expected by almost all commentators, the combinations that create meaning appear at first sight to be incompatible with the familiar grammatical categories, and there is clearly a process of reconciliation to be carried out by scholars with some generosity of outlook. At present, grammarians are extending the scope of grammar on an ad hoc basis to include some sensitivity to lexis, while lexicologists are becoming more aware of the importance of syntactic organisation in multi-word lexical units.

The key factor for future developments is the discovery that new meanings and new shades of meaning are revealed as the putative units of meaning incorporate more words — as they are seen maximally rather than minimally. This shift of emphasis reduces the role of grammar

because it restricts choice at places in structure, and it remains to be seen where it will stop.

## **HEBREW COMPUTATIONAL LINGUISTICS: PAST AND FUTURE**

**Shuly Wintner**

Computational linguistics is a research area that lies at the intersection of linguistics and computer science. Computational linguistics can be viewed in two ways: On the one hand, it is the application to linguistics of various techniques of and results from computer science, for the purpose of investigating such fundamental problems as what people know when they know a natural language, what they do when they use this knowledge, and how they acquire this knowledge in the first place. On the other hand, computational linguistics is the application of various techniques of and results from linguistics to computer science, to provide such novel products as computers that can understand everyday human speech; translate among different human languages; and otherwise interact linguistically with people in ways that suit people, rather than computers. This latter view is usually known as Natural Language Processing.

This paper focuses on natural language processing, that is, on computational applications that necessitate linguistic knowledge or emulate language capabilities. Examples of such applications include machine translation from one natural language to another; conversion of speech to text and text to speech; natural language interfaces for computational systems; automatic summarization of documents; spelling and style checking; and so forth. Advances in this field are extremely important: Natural language interaction with computational systems will make such systems more accessible, and will enable more users, including users who are not computer literate, to benefit from the advantages of the systems. Understanding and generation of speech will result in systems for automatic voice interaction and will enable hand-free (or

remote-controlled) operation of machines and systems. Speech-to-text conversion will enable hearing impaired persons to “hear” telephone conversations; text-to-speech conversion will let blind people “read” their e-mail. Automatic translation will result in huge savings, when a user’s manual of some product is written in one language only and then translated automatically to the languages of the countries in which it is sold. The field has unlimited possibilities, but the current state of the art is such that only a few of these possibilities have been actualized, and then often in a far from adequate manner.

We concentrate, in this paper, on natural language processing applications for the Hebrew language. We show that Hebrew poses additional problems for developers of programs for language processing, mainly due to its rich morphology and deficient script. We briefly review the field of computational linguistics and the problems with which it deals, emphasizing the special problems involved in processing the Hebrew language. We then survey existing systems developed for Hebrew; to the best of our knowledge, this survey covers all works published to date. Finally, we attempt to identify future needs and suggest directions for future progress.

## **MULTICULTURALISM AND MULTILINGUALISM IN ISRAEL**

**Eliezer Ben-Rafael**

Numerous cases of cleavages divide Israeli society and have an impact on language — from the Ethiopian Jews to the Circassians. The major cases discussed here confirm that multiculturalization is transforming Israeli society. It profoundly alters the nature of the project of nation-building, by recognizing a variety of interacting and intermingling constituencies. The dominant culture appears to lose its impact on the setting, at the pace of the strengthening of sociocultural groups. The question that then arises, of course, concerns the reality created by the void left by the slackening of the dominant orientation and the increasing fragmentation of the

setting. It appears, in this respect, that beyond each cleavage confronting the dominant culture by asserting its contrastive particularism, each sociocultural group is, at the same time, also significantly exposed, in a variety of manners, to that same dominant culture. The analysis evinces here the importance of the acquisition of Hebrew by all groups, although Hebrew is given different kinds of coloration and must also share its space of activity with different partners. Hebrew dwells with, and is influenced by, Yiddish and Biblical Hebrew among the ultra-orthodox, Judeo-Arabic among the Jews who emigrated from Arab countries, Russian among the Russian Jews, and Arabic among the Arabs. In addition, there is the impact of English among the privileged class. In each group, Hebrew is also granted a different meaning: It is the language of a new nation among the upper strata, a vulgar vernacular for the ultra-orthodox, the language of traditional Judaism for Israeli Jews from Arab countries, the language of a target society for the Russian Jews, and a second language for the Arabs. Yet, it remains that the generalized use of Hebrew still means a reference to a common set of symbols and, thus, the possibility of significant communication. What now keeps sociocultural groups together, as constituents of one society, is that each group has retained its distinctiveness by selecting, altering, and forging its symbols not only in reference to itself, but also through confrontation with the dominant culture. These groups create inter-cultures and inter-languages, conveying the imprint of both the dominant culture and the original cultures. Even in this era of multiculturalism, the dominant culture is still a major ingredient of the cultural changes undergone by groups at the pace of their social insertion, when efforts to retain an allegiance to legacies are concurrent with the efforts to acquire and adapt to new codes and symbols. This is the kind of glue that helps keep together those who share a setting but who, with their own hands, attach themselves to that setting's fragmentation.

## **HEBREW AMONG THE ARABS IN ISRAEL: SOCIOLINGUISTIC ASPECTS**

**Muhammad Amara**

This article investigates the main sociolinguistic aspects of the Hebrew language among the Arabs in Israel. The issues examined are: Hebrew knowledge and use, integration and diffusion of lexical items into Palestinian Arabic in Israel, Hebrew in the language landscape of the Arab villages and cities, and attitudes toward Hebrew.

The study shows that Hebrew is the main foreign source of linguistic innovation. If we compare the slow acculturation of Arab society to the culture, as relates to English, with the fast and dynamic acculturation to the Israeli Jewish culture, we obtain insights into the processes of modernization: Hebrew is a major source of modernization for the Arab society in Israel.

The prestige of Hebrew is related to the progress of Israel in many domains. Many Arabs perceive Israel as a modern country with an advanced technology. To join this progress, many young Arabs learn Israeli patterns of behavior. Despite this, Arabs attach different values to the two languages, Arabic and Hebrew. Arabs are aware that Arabic is a beautiful, rich, and prestigious language; Israeli Arabs are also aware that the mastery of Hebrew is a means for achieving economic, educational, and social levels similar to those existing among Jews. This implies that Arabs learn Hebrew for practical and integrative reasons. This situation, in fact, reflects the nature of the relationship between the Palestinians and Jews in Israel from various perspectives. First, Israel is defined and perceived by Palestinians as a Jewish-Zionist state and not a country for all its citizens. Consequently, the Palestinians are seeking to enhance their national identity in the Jewish State; Arabic serves as an important component in this national identity. Second, the Arab-Israeli conflict has not contributed to a softening of the differences between the Arab minority and Jewish majority in Israel; in some cases, the conflict

has strengthened the differences. Third, because Arabs and Jews live in separate locations within Israel, there has not been extensive contact between the two groups; this segregation has even helped preserve some distance between the two peoples. All these factors have served to deny a social convergence at a higher level toward the dominant Jewish culture and its language (Hebrew) among the Arabs in Israel.

The dual identity (Palestinian and Israeli) is reflected in the repertoire of the Palestinians in Israel. The tension between the two identities has limited the degree of convergence to Hebrew, the language of the dominant culture. This means that the Arabs adopt the strategy of linguistic integration (not assimilation). On the one hand, Israeli Arabs attempt, through achieving a high linguistic competence in Hebrew, to join the wide social network shaped by the culture of the majority; on the other hand, they preserve their identity through their mother tongue, Arabic.

### ***THE CORPUS OF SPOKEN PALESTINIAN ARABIC (COSPA)***

**Otto Jastrow**

*The Corpus of Spoken Palestinian Arabic (CoSPA)* is an ongoing project that is proceeding according to regions and that will eventually cover the whole linguistic area of Palestinian Arabic. The scope of *CoSPA* encompasses both Arabic dialectology and corpus linguistics.

*First stage: Fieldwork in the Galilee (1996–98)*

*CoSPA* grew out of a joint project of the University of Haifa/Israel (Rafi Talmon) and the University of Heidelberg/Germany (Otto Jastrow, Peter Behnstedt),<sup>1</sup> entitled “A systematic Survey of the Arabic Dialects in

1 Neither Jastrow, who has since become a professor in the University of Erlangen-Nürnberg, nor Behnstedt are affiliated with the University of Heidelberg. Werner Arnold, who is going to participate in Stage 3 of *CoSPA*, is affiliated with the University of Heidelberg.



Northern Israel.” The project, which was founded by GIF,<sup>2</sup> was undertaken during the years 1996–1998. The project covered more than 120 Arabic-speaking localities in the Lower and Upper Galilee, and we obtained first-hand linguistic data by means of questionnaires and tape recordings. The total number of informants consulted (either interviewed or recorded or both) exceeds 700. The recordings comprise more than 400 cassettes, with a total of 200 hours of recorded speech.

In addition, data were collected for approximately 30 pre-1948 localities (villages no longer existing today) for which informants, who now reside in other villages in the area, could be found. Similarly, a special investigation was aimed at retrieving the original Arabic dialects spoken by Jewish communities prior to the establishment of the State of Israel. This was accomplished for Teverya (Tiberias) and Zefad (Safed), where older members of the local Jewish community were recorded and interviewed. The city of Haifa posed an even greater challenge, because prior to 1948, the Jewish, Christian, and Muslim communities of the city spoke specific Arabic dialects. For his doctoral dissertation, Aharon Geva-Kleinberger was able to find and record a number of elderly speakers of Arabic (who had been living in the city before 1948) from all three communities.

*Second stage: Fieldwork in the Muthallath area (1999)*

In a second stage, fieldwork was extended to cover the Arabic dialects of the so-called “Triangle” (in Arabic *Muthallath*, in Hebrew *Meshulash*). This is an area that is situated south of the Galilee and that comprises approximately 25 localities, from Imm ilFahim in the north to Kufir Kasim in the south.

In 1999, the present writer completed two fieldwork campaigns, for a total of 11 weeks, in the *Muthallath* area. For the fieldwork in the *Muthallath* area, we developed a different type of questionnaire, which concentrates on the morphology of the verb. This is because the phonology

2 GIF (short for “German-Israeli Foundation”) is an agency that promotes research undertaken jointly by German and Israeli scholars.

offers less variation in the *Muthallath* area than in the Galilee, and the verb morphology is very complicated and contains some interesting points of divergence.

In this second stage, again, coverage was near complete. In each locality visited, I completed the questionnaire (an operation usually requiring the larger part of a day), and in about every third locality, I made also tape recordings of spontaneous speech.

### *Third stage: Fieldwork in the Central Israel area*

The success of stages 1 and 2, as described above, encouraged us to envisage a continuation of our work, with the aim of covering all dialects of Palestinian Arabic spoken within the State of Israel and, ultimately, in the whole linguistic area of Palestine. As the next step, another joint German-Israeli project started in 2001, again funded by GIF. The project is planned for 3 years. The German participants are Otto Jastrow (University of Erlangen) and Werner Arnold (University of Heidelberg), the Israeli participants are Rafi Talmon and Aharon Geva-Kleinberger (University of Haifa) and Arye Levin and Simon Hopkins (The Hebrew University, Jerusalem). The area of investigation will include the Tel Aviv area (Yafo, Lod, and Ramle) and the Jerusalem area, including some nearby localities that are, at present, under Palestinian administration.

### *Future stages*

The future stages will concentrate on the areas of Palestinian Arabic outside the State of Israel, that is, the so-called “West Bank”, which is currently under the administration of the Palestinian Authority. To proceed with fieldwork on a large scale, some official agreement with the local administration will be necessary. On the individual level, however, fieldwork is already progressing.

### *Characteristics of CoSPA*

As described above, the scope of *CoSPA* encompasses both Arabic dialectology and corpus linguistics. As a multi-purpose corpus,

— It serves the needs of dialectology, providing the basis for grammatical description, dialect geography (including maps), and a lexicon of the respective areas.

— It equally provides the basis for ethnographic and folkloristic studies and allows the publication of collections of oral literature.

— It can provide the basis for computer-based research along the lines of “corpus linguistics”.

*CoSIH and CoSPA: Points of convergence*

There is a considerable overlap between *CoSIH* and *CoSPA* in terms of territory and population, because all speakers of Palestinian Arabic who are citizens of Israel are also members of the Hebrew-speaking population. There is continuous Hebrew interference in Palestinian Arabic speech and, on a smaller scale, Palestinian Arabic influence in spoken Hebrew. As *CoSIH* and *CoSPA* proceed, it will be possible, for the first time, to study the interrelation of the two languages on a sound empirical basis.

**FROM MONOLINGUAL TO MULTILINGUAL?  
EDUCATIONAL LANGUAGE POLICY IN ISRAEL**

**Elana Shohamy and Bernard Spolsky**

Although Israel is perceived as a monolingual country, the reality is that it is multilingual: The different groups residing in the country use a large number of languages. The paper describes the process and cost of the revival of spoken Hebrew that began at the end of the nineteenth century. This process continued in Palestine, where Hebrew was reinforced by a strong Zionist ideology that associated language with national unity and discouraged immigrants from maintaining home languages, even Jewish languages such as Yiddish and Ladino. The cost of the revival of Hebrew was the loss of the rich linguistic repertoire of the Jews, a traditional trilingual pattern, typical of the years in Diaspora: Hebrew and Aramaic for sacred and secular literacy; a Jewish language such as Yiddish or

Ladino for oral functions within the community, and a territorial gentile language for external contacts. The monolingual Hebrew ideology was bolstered by myths and assumptions that encouraged a new language policy, based on the following assumptions: (a) Hebrew will be learned by immigrants only if all home languages are abandoned; (b) acquiring Hebrew will lead to national unity; and (c) knowledge of Hebrew is key to acculturation and integration, which will be slowed if immigrant languages and memories of Diaspora continued to exist. The result has been a monolingual society with strong dominance of Hebrew, loss of Jewish and other heritage languages, and general decrease of national language capacity.

In practice, monolingualism has been more of an ideology than a reality, as a highly complex pattern of multilingualism has been created by groups living in Israel before independence, by the continuous waves of immigration, and by the large number of foreign workers. Although approximately 4,500,000 Israelis are estimated to have functional competence in Hebrew (there has been no language question on the census since 1983), other languages are represented: about 1,400,000 speakers of Arabic, 800,000 of Russian, 200,000 of French, 215,000 of Yiddish, 250,000 of Rumanian, 100,000 of Spanish, 60,000 of Hungarian, 60,000 of Persian, 60,000 of Amharic, and 50,000 of other languages of the former Soviet Union and other languages used by the approximately 200,000 foreign workers (see Spolsky and Shohamy, *The Languages of Israel: Policy, Ideology and Practice*, 1999, for a thorough discussion of the topic). These languages create pockets of resistance to the monolingual ideology. Thus, Hebrew is now being challenged by Russian, Arabic, Yiddish, and the languages of the foreign workers, and because Hebrew is forced to compete with English in an increasing number of domains. As well as widespread plurilingualism, the beginning of respect for diversity is to be seen in signs of multilingual ideology. This was behind the new language in education policy for Israeli schools, adopted by the Ministry of Education in 1996. This was the first official document stating that each group, Jews and Arabs, should learn three languages and should be

encouraged to learn an additional home, community, or world language as well. The multilingual policy is based on the assumption that Israel consists of a large number of ethnolinguistic groups, that the languages of immigrants and indigenous groups are assets and not liabilities, that language learning should be additive rather than subtractive, that plurilingualism is part of multiculturalism, and that different languages are needed for different purposes. For Jews, the languages are Hebrew, Arabic, and English, plus a heritage/community/world language; for Arabs, it is Arabic, Hebrew, English, and additional languages. The policy encourages immigrants to maintain home languages. The publication of the new language policy is encouraging, as it does not view other languages as a threat to Hebrew. However, there remain serious questions about implementation of the policy, as result of complex pressures from a variety of directions. One is the lack of motivation to learn additional languages, especially in the case of learning Arabic, given the political conflict. Another is the reluctance to teach spoken Arabic, and there are some who fear that English in Israel poses a threat to Hebrew. In addition, it is not clear that community efforts to maintain Russian will succeed. In sum, the new policy offers a guide on how to develop a successfully multilingual society, but there are no guarantees that the policy will be implemented by governments that have more urgent problems.

## **LINGUISTIC VARIATION ACROSS GENERATIONS IN ISRAEL**

**Yaakov Bentolila**

Israeli Hebrew comprises roughly two broad linguistic variants: “General” (GH) and “Mizrahi” (MH). The former enjoys social prestige and is also the most widespread by the elite. It is, therefore, commonly spoken by a constantly growing part of the population. The “Mizrahi” variety reflects Hebrew as it was pronounced in Arabic-speaking countries. Immigrants from those countries form the lower ranks of Israeli society. Mizrahi Hebrew is socially marked. Linguistically, it is characterized,

*inter alia*, by the full phonetic realization of the pharyngeal fricatives: the sociolinguistic variables (ħ) [HET] and (ʕ) ('AYIN).

The emancipation of Mizrahi Jews in Israel is an ongoing social trend with cultural and political aspects. Culture-oriented Mizrahi leaders and intellectuals might be motivated by the will to gain recognition or esteem for the values of their ethnic identity, and those include the MH pronunciation, especially HET and 'AYIN. On the other hand, young Mizrahi activists may find themselves in a conflicting situation. Their efforts to integrate with the leading elements in the society compel them to use the prestigious GH norm. The situation becomes a knotty one when ethno-cultural arguments are used for political purposes; thus, the political context, when devoid of ethnic concerns, promotes the use of GH, while the cultural or the ethnic contexts, either with or without a political bias, tend towards MH. Also, on some occasions, Mizrahi persons, who otherwise speak in the MH variety, may use intentionally the GH pronunciation in interactions with persons of their own community; in exhibiting linguistic manners of the influential elements of Israeli society, they seek social status.

From what is commonly known about MH and its users, the phoneme /ħ/ is considered as a sociolinguistic indicator, that is, as a variant that shows social or ethnic distribution. It might be interesting to regard it as a sociolinguistic marker, that is, one displaying stylistic distribution. I found that, in Romema, (ħ) was still largely an indicator among the elderly (males or females), among the majority of the younger males, and among females with poor educated background. However, it had begun to function as a marker among more educated young females and among young males involved in political activity.

I think that a sociolinguistic indicator becomes a marker via gradual change and is characterized by a growing awareness of its social values by part of the population, for example, younger females and politically active young males. That gradation from no variation at all (a status of "indicator") to an evolving consciousness of the social value of a variant (a status of "marker"), promotes the spread of variation and is

more readily observable in small communities that are characterized by a close-knit social network.

As far as Romema is concerned, sex, age, socio-political involvement, and outward connections are most significant in the linguistic choices made by the speakers. Although systematic research has not yet been undertaken in Israel on that specific subject, one cannot help noticing that the linguistic behavior of children is often very dissimilar to their parents' linguistic behavior, mostly as the result of different social routes at adult ages and of ties with peer groups in the adolescent phase of life. In constructing a corpus, I believe we should insist on gathering data not only from individuals, but also from parents and close kin, so that the relevance of family ties to linguistic variation and change can be examined.

## **LANGUAGE VARIETIES IN CONTEMPORARY HEBREW**

**Ora (Rodrigue) Schwarzwald**

The paper is an overall survey both of the linguistic varieties found in Modern Hebrew in Israel and an update review of the research on each linguistic variety. The classification of the linguistic varieties is based primarily on text typology, as described by external and internal criteria by Sinclair, Ball and their team (EAGLES96), in which text refers to any linguistic performance, oral or written. The external criteria for text typology include (a) origin, (b) state, and (c) aim of a given text. The internal criteria for text typology include (a) topic and (b) style. The external criteria determine the internal type of text.

The most significant factor in determining language varieties in Modern Hebrew is the text origin, namely, the speakers' features, their locations, times of speech production, and the roles of the participants. The speakers' features include the following types: (1) ethnic origin, (2) socioeconomic status, (3) gender, (4) religiosity, and (5) age. The state of the text refers to whether it is oral or written. Oral texts vary according to the speech pace and cautiousness. Written texts range from formal to informal, literary to

non-literary writings, printed and non-printed, and so on. The aim of the text refers both to the target audience of the text and to the producer's intentions in producing the text.

The internal text criteria are interwoven among the external ones. Therefore, the most significant varieties are content oriented, depending on the subject matter, ideological attitudes, and social desirability. The style varies between formal and informal types.

The research survey reveals that the ethnic varieties of speech were studied in the 1950s, and most of the research done in other sociolinguistic areas of Modern Hebrew started in the 1970s. Research in gender and religion linguistic varieties is the most recent research. Based on pure linguistic observations, the paper also raises many issues of language varieties that ought to be further studied scientifically.

## ON LANGUAGE, LAW, AND SOCIAL JUSTICE

Zohar Livnat

The various areas in which language and the law overlap have, in recent years, attracted the attention of researchers from different disciplines. The common areas in which the linguist can contribute actively and substantially to the justice system are gradually becoming more clear. Linguistic tools may serve as aids in the legal decision-making process in the courts or in judicial review. This paper will shed light on those points that may be especially important and productive in this context.

### *Legal register, social justice, and the demand for the use of Plain Language*

The linguistic debate surrounding the legal-judicial register need not be restricted to a mere description. This area of discussion can be expanded to include the unique difficulties that the legal register presents to the public and can even suggest solutions to this problem.

The unique properties of the legal register make legal texts difficult for the uninitiated to understand. On the assumption that the rights of citizens



(as well as those of residents or tourists) are at least partially dependent on their ability to comprehend the texts in which these rights are described, and on the assumption that involvement in a legal process of any kind is always contingent on contact with spoken and/or written texts, it is understood that our interest here is with issues related to inferiority or inequality before the law.

Non-native speakers are likely to require interpretation during any contact they may have with the legal system. However, the courts tend to keep the use of professional interpreters to a minimum, calling upon them only when left with no other alternative when conducting a trial. An interrogation conducted in a language in which the defendant is not fluent has far-reaching implications for the manner in which the testimony is received. Despite this, Israeli law does not go into sufficient detail on the question of when a defendant “does not know Hebrew” and who is qualified to interpret for the defendant in such a case.

The legal register may be difficult for native speakers to understand as well. Recognition of this triggered the demand to simplify legal language a number of decades ago. This recognition also led to the establishment of movements promoting the use of Plain Language in legal texts in many places in the world. In Israel, such voices are not being heard. In my view, the possibility should not be ruled out that — knowingly or unknowingly — the legal community has perpetuated its public and economic standing by means of a linguistic style that is partially or wholly incomprehensible to large sections of the public.

*“Legal linguistics”: Linguists as expert witnesses in court*

In many Western countries, there has been a tendency, in recent years, to summon linguists to testify as expert witnesses on linguistic matters relevant to the court debate. Most cases involve two types of testimonies — those related to the ability of a particular addresser to produce a certain text (e.g., in matters related to the determination of the authenticity of a confession), and those related to the ability of a particular addressee to understand a certain spoken or written text.

The opinion provided by an expert linguist may also aid in legal interpretation. To illustrate the possible contribution of linguistic expertise in this matter, this paper discusses a case brought before an Israeli court, in which a considerable portion of the debate focused on the meaning of the word “violence” and on the question of whether “verbal violence” is indeed a form of violence. A linguistic analysis proves that the legal debate, described by the judge as “language-linguistic”, does not always conform to the analysis as seen through linguistic eyes and may even lead to an opposite conclusion.

Linguists can place themselves at the disposal of the judicial system as experts, show the places where linguistic expertise is necessary, and propose linguistic tools from various fields, such as second-language acquisition or reading comprehension, to aid in situations related to the judicial mechanism.

*Rhetoric and the legal discourse: Form versus content*

The debate on rhetoric in the legal context can be a critical one, for example, in the analysis of testimonies and the designation of the situations in which justice may not necessarily have been served. Testimonies given in court may often be alternative versions of the “truth”, and the judge must decide “whom to believe”. This choice may be affected not only by the content of the testimonies, but also by their form or even phrasing. False testimony presented articulately, fluently, and confidently may make a better impression than a truthful one presented haltingly, irresolutely, and timorously.

## **MODERN HEBREW AND ITS CLASSICAL BACKGROUND**

**Moshe Bar-Asher**

This article investigates the history of modern written and spoken Hebrew. We may consider the middle of the eighteenth century, when Moses Mendelsson published *The Kohelet Musar* journal (c. 1755), the starting

point of modern written Hebrew. The arrival of Ben-Yehudah in Eretz Israel in 1881 may be taken as the starting point of spoken modern Hebrew. Clearly, modern Hebrew maintains the historical continuity of the Hebrew language. Most of the lexical and grammatical elements incorporated into modern Hebrew have been drawn from the Hebrew of ancient periods.

Nevertheless, it is clear that many syntactic neologisms have been derived from earlier strata. In other words, modern Hebrew makes diverse and innovative usage of the elements it adopts from classical Hebrew, which I demonstrate through two grammatical patterns.

### (1) *The pa'al pattern*

This pattern is found in classical Hebrew (Biblical Hebrew and Mishnaic Hebrew), for example: **ירקרק, אדמדם**. Even in ancient times it was not clear whether **אדמדם** meant red, very red, or slightly red. Indeed, sources from the tannaitic period provide evidence of all three opinions. In modern Hebrew, it was decided to adopt the *pa'al* pattern to indicate the diminutive. Hence, **כלבלב** denotes a small dog, and **קטנטן** means very small.

### (2) *The pa'il pattern*

In classical Hebrew, we find *pa'il*, which functions as a verbal noun or indicates a season of the year, for instance, **חריש** (the activity of the plow or the plowing season). We also find *pa'il*, which is the passive participle of the *pa'al* pattern, that is to say, a variant of *pa'ul*: **קריא** like **שליח**, **קרוא** like **שלוח** are passive participle forms. However, in modern Hebrew, the forms **אכיל**, **קריא** and so forth, serve as adjectives of the potentiality. Therefore, **קריא** means readable, and **אכיל** means eatable. This innovation was apparently inspired by the modern European languages, which create this kind of adjective. In summary, we can determine that written and spoken modern Hebrew often adopts classical elements from the rich tradition of classical (Biblical and Mishnaic) Hebrew and makes new usage of them. What I present here is but a brief demonstration of this trend.

## **THE EMERGENCE OF SPOKEN ISRAELI HEBREW**

**Shlomo Izre'el**

Formerly used mainly as a literary and liturgical language, Hebrew was transformed at the turn of the twentieth century into a full-fledged, vernacular language and the national language of the Jews in Israel. The term "revival" for the emergence of Hebrew in Palestine cannot be justified. Hebrew never died, and the language served for almost two millennia, not only as a written medium, but also for oral communication between Jews wherever necessary.

The written language was never frozen or rigid. It was constantly changing, and influences from internal Hebrew strata (especially from local vernaculars — languages spoken by Jews wherever Hebrew was written) have always made their impact on Hebrew. Regressions to more "original" or "pure" Hebrew, as was the case during the time of the Enlightenment, are rare and must be regarded as partial.

The course taken by written Hebrew toward its use in Palestine and Israel as the main medium of written communication must be regarded as sequential and gradual to various degrees. This is obviously not the case with the emergence of spoken Hebrew. When massive waves of immigration from many countries began to reach Palestine, a multilingual society with a pressing need for immediate communication began forming. It was at this point that the transformation of Hebrew took place most rapidly.

The result of both processes, the sequential development of written Hebrew and the rapid regeneration of spoken Hebrew, is the contemporary linguistic continuum of the Hebrew language, as it is used in Israel. From the linguistic point of view, Modern Hebrew, as compared with earlier stages of Hebrew, has a significantly different structure.

Shaped mainly under the influence of European strata at the beginning of the twentieth century, one may compare the emergence of Israeli Hebrew to the emergence of Creole languages. This paper describes

similarities and dissimilarities between the emergence of Hebrew and the creolization process and tries to show some advantages in setting the recent history of Hebrew under this alternative perspective.

To understand the active processes within the creation of Modern Hebrew, research should take the following directions:

1. Search for evidence of spoken Hebrew in the early years of its emergence,
2. Search for evidence of prior creolization in data from contemporary Israeli Hebrew, and
3. A reevaluation of the sociolinguistic and, especially, previous linguistic research on the emergence of twentieth-century Hebrew.

All aspects raised in this paper or implied from its subject matter must and can be investigated: the extent to which contemporary Hebrew is different from all previous layers of the language; the gap between spoken and written varieties of Hebrew; the interrelationship among the continua of varieties within each and the impact of any of the existing varieties on each other; the history of Modern Hebrew on both its written and spoken continua; how Hebrew fits into the larger continuum of contact-induced languages; and how the emergence of Hebrew is related to creolization. All of these issues and many other related and unrelated questions must and can be investigated. However, investigations such as these can occur only where data exist. Without the compilation of an Israeli Hebrew corpus, none of these issues will emerge from obscurity. In fact, none of these topics can be investigated at all, unless we have at our disposal a corpus of Israeli Hebrew, both written and spoken.

## **PHONOLOGICAL AND MORPHOLOGICAL VARIATION IN SPOKEN HEBREW**

**Shmuel Bolozky**

The article starts with a discussion of some of the processes responsible for phonetic variation in Israeli Hebrew, mostly as a function of rate of

speech and degree of attention. Of the different vowel elision processes, the most common involves the reduction of *e*. The “minimal” vowel *e* may be elided when it is no longer required to correct violations in the sonority scale. Thus, for instance, based on *kélev* “dog” ~ *kla-vím* “dogs”, one would expect *yéled* “child” ~ *yla-dim* “children”, but inasmuch as *yla* violates the sonority hierarchy, the actual form is *ye-la-dim*. However, if a vowel precedes, as part of an appended proclitic, the syllable structure may be modified so that *y* is post-vocalic, which removes the need for this *e*, as in *a-yè-la-dím* ~ *áy-la-dím*. *e* is also likely to be elided in affixes in casual/fast speech. Affixes are more amenable to reduction than stems, because of the ready accessibility of affixes and their high frequency, which makes them easily recoverable in context. Thus, *u mèlaméd ivrít* “he teaches Hebrew” ~ *ùmlamédivrít*. Occasionally, *e* may be deleted in stems as well, provided that they are frequent enough, as in *raíta kvar et a-séret a-ze?* “have you already seen this movie?” ~ *raítakvartasértazè*. In rapid/casual speech, *e* may be elided even when it is normally maintained to prevent the occurrence of geminates: *od ló raíti zalelán kazè* “I’ve never seen such a glutton” ~ *odlóraítizallánkazè*. An unstressed *e* is often assimilated into an immediately following unstressed vowel that has resulted from the loss of a glottal or pharyngeal consonant, as in */šè'oním /* “watches” > *šèoním* > *šdoním* > *šo:ním* (> *šoním*), *lèasbír* “to explain” > *làasbír* > *lasbír*. *i* may be elided in casual/fast speech as well, most commonly also in affixes, when future forms functioning as imperatives are reduced, as in *tíškeví bešéket* “lie (down) f.s. quietly” ~ *išekví bešéket* ~ *čekví bešéket* ~ *škevíbešéket*. The other three vowels, *u*, *o*, and *a*, are generally more resistant to casual deletion, except for some function words or clitics and a few frequent items. *a* is unaffected even by (substandard) *hif'il* centralization of prefixal *i* to *e*, as in *hisbír* “he explained” > *hesbír* vs. *masbír* “explain” > \**mesbír*. On the other hand, it is claimed that variations like *hipil* “he dropped (tr.)” ~ *hepil* are not due to centralization, but rather to analogy with the large class of the “hollow” *heCiC* type. This article then shows that vowel reduction/elision and assimilation observed in Israeli Hebrew appear to have had precedents in

Biblical Hebrew, as in Gen 38:27 *tə'ōmīm* “twins” ~ Gen 25:24 *tōmīm*, IKgs 18:12 *ləhaggīd* “to tell” ~ IIKgs 9:15 *laggīd*.

The next section describes consonant reduction. In general, sonorant consonants are more amenable to elision in casual and/or fast speech — again, in direct relationship to the frequency of the morpheme concerned. The likelihood of reduction is (possibly) dependent upon the status of the sonorant consonants on the sonority scale, because syllables in which there is a small sonority contrast between onset and peak are not optimal. The sonorant *y* is, thus, the most likely consonant to delete, particularly in a *yi* sequence, as in *yisraél* “Israel” > *israél*, but also with other vowels, as in *yèladīm* “children” > *èladīm*. The next likeliest sonorant consonant to reduce is *l*, particularly in grammatical words, as in *à-xavér šelxà* “your m.s. friend” > *àxavér šexà*; and so on. Consonant assimilation and merger processes are subsequently discussed: (anticipatory) voicing assimilation, as in *tizkór* “(you will) remember” ~ *tiskór*, *tisgór* “(you will) close” ~ *tizgór*; assimilation of a dental nasal to a velar one before a velar stop, as in *pinkás* “notebook” ~ *piŋkás*; geminates occasionally, arising across morpheme boundaries in casual/fast speech and their possible simplification, as in *avád+ti* “I worked” > *avádeti* ~ *avátti* ~ *aváti*, *yašán+nu* “we slept” ~ *yašánu*. Next is a description of variations such as *anì avó* “I’ll come” > *anì yavó* ~ *nyavó* ~ *navó* in the first person singular future tense, followed by a discussion of the realization of the glottal stop (or other segments merged with it), the claim being that when a glottal stop does surface in Israeli Hebrew, it is only as an onglide to a heavily stressed vowel, usually for the purpose of contrastive emphasis, as in *amárti kar’á*, *ló kará* “I said ‘she read’, not ‘(she’s) cold’.” The following section tries to account for all occurrences of penultimate stress in Israeli Hebrew, demonstrating that it is more common and more “natural” than generally assumed. It is followed by description of secondary stress, which is shown to generally fall on every other syllable, counting backward from the main stress (e.g., *vèaxàverá* “and the girlfriend”), except for two-syllable troughs resulting from stress clash (e.g., *šlošá yèladīm* “three children” > *šlošá yeladīm* ~ *šlošá yladīm*), or

stress shift caused by the same (e.g., *atá bá* “are you coming?” ~ *àta bá* ~ *tabá*). The survey of phonological alternation concludes with an introduction of a sort of pre-tonal lengthening, which may be observed occasionally in Israeli Hebrew, as in *ze šo:né* “it’s different”, *ma:hír meód* “very fast”.

In general, phonological variation is characterized by the following: assimilation, reduction, and deletion are directly related to frequency of use and (related) likelihood of recoverability. Consequently, affixes and prosthetic vowels are the most likely to be affected. The likelihood of vowel reduction or elision is essentially in inverse relationship to the degree of sonority. The opposite applies to consonants, because of a preference for maximal sonority difference between *C* and *V* in a *CV* sequence in a syllable. Minimal sonority difference between a consonant and a vowel in a *CV* syllable facilitates elision of *C*. Consonant assimilation is usually regressive and is strongest if the two consonants concerned are tautosyllabic. Other assimilations and simplifications may be explained by “ease of pronunciation”.

It is more difficult to speak of variability in the same sense as in phonology, when the morphological component of Israeli Hebrew is concerned. In morphology, there is no true variation, in that each realization is at least minimally different semantically from its apparently equivalent alternative. Nevertheless, tendencies in recent innovation, formulated through dictionary comparison or the study of *hapax legomena* in corpora (cf. Bolozky, *Measuring Productivity in Word Formation: the Case of Israeli Hebrew*, 1999) do point to certain interesting variations in the realization of certain semantic features, some of which can be attributed to register. For instance, causative verbs tend to be realized in *hif'il* in the literary register, but in *pi'el* in the colloquial and in any non-literary register, for example, *he'egid* “form a concern” vs. *fideax* “caused embarrassment”. When a verb is involved (in the base, or in the definition of the target meaning), adjectives are realized in the *meCuCaC* pattern; otherwise, the default adjectival realization is with the suffix *+i*, for example, *medupras* “depressed” (cf. *dipres* “depress”, *diprésya* “depression”) vs. *bixuti* “of



safety” (cf. *b(e)tixut* “safety”). Recent formations for agent nouns may be realized, either discontinuously, in the *CaCCan* pattern, or linearly, by adding *+an*. There is no significant functional difference between forms belonging to the two types. Examples: *kanyan* “buyer, purchaser” (cf. *kana* “buy, purchase”) vs. *mis’adan* “restaurant proprietor” (cf. *mis’ada* “restaurant”). In the colloquial, instruments that in normative Hebrew are realized in *maCCeC* merge with the agentive *meCaCeC* pattern, forming a type of “performer” category, covering both agents and instruments performing an action, as in *maghec* “iron” *megahec* for “one who irons” — and for “iron” as well. Also in the colloquial, locatives that in normative Hebrew are realized in *miCCaCa* merge with the instrumental *maCCeCa* pattern, possibly because of the orthographic similarity, because *maCCeC(a)* is more transparent/specialized/prominent semantically and because there is a preference for the more prominent vowel *a*, for example, *mixbasa* “laundry (room)” > *maxbasa*. A distinction is emerging between an *+er* suffix with penultimate stress, originally from Yiddish/English with agentive meaning and now mostly instrumental — and *+onér* or *+yonér*, with final stress, originally from French (or German), which is a fairly productive agentive pattern today. It appears that the two patterns are distinguished, primarily, by the location of stress. Illustration: *kvéčer* “stapler”, from *kvéč* “squeeze (from Yiddish)”, vs. *krizyonér* “one easily going into rage”, from *kríza* “rage”. Variation in the stress patterns of gentilic names and residents of geographical locations is claimed to be caused, at least in part, by assignment of either *+i* with stress on the preceding syllable or *+aí* (with final stress), for example, *albánya* “Albania” *+i* > *albáni* “Albanian” vs. *petaxtikva* “Petah Tikva” *+aí* > *petaxtikvaí* “resident of Petah Tikva”. Generally, variation in the morphological component is manifest, primarily, in the choice of different morphological patterns (or the location for stress assignment) to capture closely related semantic notions. Sometimes, the semantic target is the same, and the different patterns are chosen on the basis of speech style alone (literary/everyday, formal/casual, etc.). In other cases, the choice is based on some inherent distinctions such as syntactic category (usually verb vs. non-verb), or closely related semantic fields (e.g., agent/instrument).

## THE STUDY OF MODERN HEBREW SYNTAX

Geoffrey Khan

This paper presents a brief survey of previous work on Modern Hebrew syntax.

### *1. Debate over linguistic system, object of study*

The structure of the Modern Hebrew language that is used today emerged at the time of its formation at the beginning of the twentieth century by a combination of features that are characteristic of different historical layers of the language. In very broad terms, the morphology is based on that of Biblical Hebrew whereas the syntax is closer to that of post-Biblical Rabbinic Hebrew (cf. Sáenz-Badillos, *A History of the Hebrew Language*, 1993). The language was also exposed to external linguistic influence, mainly from European languages, in its lexicon and also, to some extent, in its syntax. In the first decades of its existence, the attention directed by scholars and teachers to the grammar of the language, in both its written and spoken form, had almost exclusively a prescriptive rather than a descriptive purpose. There was particular reluctance to make an objective description of the grammatical structure of the spoken language. It was not until the 1950's that such synchronic structural analyses began to be made.

### *2. Types of analysis*

#### *Structuralist*

In the 1950's the tension between the normativists and linguists advocating synchronic descriptivism of linguistic reality came to a head. One of the most energetic proponents of objective linguistic description was Haiim Rosén. While recognizing that every speaker had an idiolect, he held to the view that the language of the speech community as a whole had a unitary system. Rosén applied himself to a structuralist study of syntax in his book *Ivrit Tova* (1958). In the 1950's and sixties some descriptive studies of various syntactic constructions in Modern Hebrew were made

by other scholars. Rosén's structuralist work on syntax influenced several subsequent studies. Most structuralist approaches worked with the notion of single language system and did not incorporate into their framework differences of register and sociolect.

### *Generative transformational syntax*

From the early sixties Chomskian transformational syntax influenced work on Modern Hebrew syntax. Some of the early transformational analyses were based on the theoretical model that was presented by Chomsky in his first work *Syntactic Structures* (1957), which treated only a small range of constructions. In the late sixties and in the seventies a number of similar studies were made of other areas of Modern Hebrew syntax, broadly in the framework of Chomsky's 1965 model. Some, but by no means all, of these treatments of Modern Hebrew syntax in the framework of generative grammar have contributed new descriptive facts about the language. In generative approaches, however, the object of study was regarded as a single grammatical system. All types and strata of Hebrew are, therefore, frequently fused together and separate descriptions are not regarded as being necessary.

### *Information structure, discourse and pragmatics*

Since the 1970's a number of studies of Modern Hebrew syntax have gone beyond an analysis of structure and examined the function of various syntactic constructions within the context of their use. This has involved investigating the information structure underlying the syntax and dealing with concepts such as given and new information and contrastive focus. Some studies have taken into account other features of the discourse in which the constructions occur. In the case of spoken, conversational language, syntactic constructions have been elucidated by studying the pragmatic context of their use.

### *Typological and diachronic*

We may classify some work on Modern Hebrew syntax that has been

published in the last few decades broadly as typological and diachronic. These studies examine the relationship between certain Modern Hebrew syntactic structures and those in other languages of the same type, or the relationship between Modern Hebrew structures and those found in earlier historical layers of Hebrew. They are generally, therefore, comparative or contrastive in nature. A number of studies have pointed out similarities between Modern Hebrew syntactic constructions and those of other modern languages. Many of these explain the similarities as the result of linguistic influence on Modern Hebrew rather than simply typological parallelism. We may include in this section the numerous studies of syntactic constructions that have newly emerged by a natural process of linguistic change in some varieties of spoken Modern Hebrew and deviate from what is found in earlier layers of the language.

### ***3. Distinction between varieties of speech***

Most work on Modern Hebrew syntax so far has not made a clear distinction between different layers of language, but rather focused on standard usage in writing and speech, or on non-standard features of spoken language without specifying anything concerning the background of the speakers. It is now recognised that Modern Hebrew exists not only in a variety of registers, but also in a variety of sociolects. There have been a few studies in recent years that devote attention to syntactic differences in forms of Hebrew spoken by various clearly defined socio-economic layers of society.

Some work has been done on the syntax of children's speech. There have been studies of what is defined as slang, which includes some syntactic features (e.g. Sappan, *Darxe ha-slang* (1963)).

### ***4. Corpus based study of syntax***

A few remarks concerning the treatment of syntax in *The Corpus of Spoken Israeli Hebrew (CoSIH)* and the possible areas where a corpus based study of syntax could be of particular value are as follows. In the preparation of the corpus, it is important at the outset to represent in the transcription

of the texts as much information as possible that is of relevance for the interpretation of syntax. In the case of spoken forms of the language this is more than a straight representation of syntactic structure. It is crucial, in my opinion, to represent also suprasegmental features, particularly the boundaries of intonation groups and the position of the nuclear stress in the intonation group. In casual speech, the placement of intonation group boundaries and the positioning of the nuclear stress are important signals that are integrated with syntactic structure. Such suprasegmental features are likely also to vary in different types of speech and across speakers of different backgrounds. Since one of the major aims of corpus research is to reveal distinctions of this kind in speech, it is clearly essential that the transcription of texts in a corpus of spoken Hebrew should include the notation of such features.

### **THE SOCIOLINGUISTIC PHENOMENON OF V FORM IN HEBREW IN THE BRITISH MANDATE PERIOD**

**Yael Reshef**

The paper discusses the employment of formal terms of address (V form) in the early layer of Modern Hebrew. An examination of administrative correspondence of the Tel Aviv municipality, in the first decades of its existence, revealed that during that period, V form is used regularly. The regular mode of address in such correspondence is third person singular for a single addressee, and second person plural for a group of addressees. The first signs of a gradual process of substitution of V form by the second person singular appear in the 1930s. However, V form remains the dominant mode of address in administrative correspondence throughout the 1940s. Only in the 1950s, does second person singular become habitual, alongside V form in this register.

The employment of V form reflects the transference of linguistic habits from the substrate languages of the first generation of Modern Hebrew speakers. The choice of the third person singular as the Hebrew realization

of this phenomenon is based on the heritage of previous linguistic layers. Such usage is occasionally attested as a mode of address in Biblical Hebrew, and a similar practice remained operative in written Hebrew until the modern period. The paper discusses the optional status of the phenomenon in the period's written language and provides evidence for the phenomenon's partial distribution in the period's spoken language as well. The data indicate that a general and regular realization of V form is unique to administrative correspondence and singles out this form within the period's linguistic system. The regular differentiation between administrative correspondence and other manifestations of the period's language attests to the existence of register differentiation in Modern Hebrew at that early stage in its development.

The gradual exclusion of V form from Modern Hebrew fits in the sociolinguistic context that influenced the development of the linguistic system. Cross-linguistic studies show that Modern Hebrew is characterized by a preference for direct modes of expression. This preference, reflecting the egalitarian ethos that dominated Israeli society in the formative *Yishuv* period, developed gradually and became characteristic of Israeli culture in the 1950s. The disappearance of V form, which occurred during that period, is one manifestation of the sociolinguistic processes that shaped Modern Hebrew, according to the values of the society it served.

## **THE SIMPLE IMPERSONAL CONSTRUCTION IN TEXTS REPRESENTED AS COLLOQUIAL HEBREW**

**Ron Kuzar**

The paper analyzes the *simple impersonal construction*, which is typical of colloquial Hebrew, although it is making headway into the written language as well. The construction is built around an indeclinable predicate or an invariable predicate in the third person, masculine, singular

form. The unmarked arrangement of this construction is rhematic, that is, the predicate appears in initial position:

- 1) *keday* (indecl.) *lexa et hamixnasayim ha'ele* (m.pl.)  
 worthwhile to.you OBJ the.pants these  
 "You'd better go for these pants."

Predicates that participate in this construction belong to two major semantic groups: modals and existentials, each comprising a number of sub-groups. There is no clear-cut semantic division between the groups and the sub-groups. Thus, predicates may be in more than one group, or they may have contextually contingent meanings within the total range of modal and existential meanings, forming "family relatedness", in Wittgenstein's sense.

A predicate in this construction may require two arguments: an optional object, which represents the relevant person, and a second argument, which represents the entity whose existence is declared or about which judgment is passed. The syntactic status of the second argument is at stake in the following discussion.

The two most obvious properties of subject and object are subject-predicate agreement and object marking (through case or preposition), respectively. In written/formal style, most of the predicates of this construction are in agreement with the second argument, that is, the latter is positively marked as subject, and others are clearly marked as objects. In colloquial language, the status of the second argument is unclear. In some cases, as in the example above, the second argument is clearly an object, because it is preceded by the overt object marker *et*. Such is the situation when the object is indirect as well:

- 2) *od lo nim'as lexa mehabdixa hazot?*  
 yet not fed.up (m.sg.) to.you from.the.joke this (m.pl.)  
 "Aren't you fed up with this joke yet?"

In many cases, however, the second argument lacks both agreement with the predicate and any object marking:

- 3) *davka mat'im lexa hamadim*  
actually suits (m.sg.) to.you the.uniforms (*plurale tantum*,  
m.pl.)  
"The uniform actually suits you."
- 4) *me'anyen hakatava hazot*  
interesting (m.sg.) the.report this (f.sg.)  
"This report is interesting."

What is most typical of the second argument, then, is the lack of marking, rather than any positive display of its syntactic nature. This is, therefore, one area of Hebrew syntax in which there is a real difference between the formal and the colloquial varieties of the language.

There are about 90 documented examples in this paper. Although some examples are taken from Israeli radio, television, and newspapers, most examples come from Youval Shimoni's (1999) novel *kheder* ("A Room"). Data of this kind are problematic, because they cannot be considered real colloquial Hebrew, but, rather, Hebrew represented as colloquial in a literary text. The results of my study should, therefore, be taken as provisional and will hopefully be testable when *The Corpus of Spoken Israeli Hebrew (CoSIH)* is completed.

## **BETWEEN SPOKEN AND WRITTEN LANGUAGE: EXAMINING PARALLEL SPOKEN AND WRITTEN TEXT**

**Esther Borochofsky-Bar Aba**

The study of the spoken language, particularly that of the last 20 years, has already proven that written language is not merely a transfer of the spoken medium, and vice versa. Currently, it is widely accepted that the different circumstances under which the two media are employed, and the different discourse principles according to which they operate, cause considerable differences between the written and the spoken text.

Due to the lack of a corpus of spoken Hebrew, I investigate differences



between the written and the spoken language in a specific way: the comparison of texts that, due to social (and not linguistic) needs, have both a spoken and a written version. In this paper, I examine two such texts: stenographic recordings of legal proceedings and Hebrew subtitles on television for certain Hebrew programs. The texts are produced orally, and a need arose for their written version. As a result, it is possible to track the necessary changes that occur during the transfer. The identification of the differences that result from the change in medium enables us to receive a picture of certain types of linguistic components that can be found in spoken texts, but not in written texts, and vice versa.

It should be said that only certain properties of the spoken language are discovered in this research, because many characteristics of the spoken language can be investigated only by using large-scale, spoken corpora and not necessarily those that have parallel written texts.

This paper shows typical traits of each of the genres investigated and common traits indicating the uniqueness of the spoken language.

The language used in subtitles is characterized both by minor deletions of components that are not essential, insofar as the content conveyed is concerned, and by corrections of the spoken text to change it to a normatively superior text, while simultaneously attempting to preserve a sentence structure as close as possible to that of the spoken text. The minor changes need to be as imperceptible as possible, on the one hand, to allow the viewers who are hard of hearing to feel the vitality of the language, but, on the other hand, to avoid interfering with the viewers' fluent perception.

The written protocols of legal proceedings (as opposed to the things said in the recorded texts) omit many minor technical utterances that are not part of the legal proceedings. The written version presents minor changes similar to those in subtitles, but there is a tendency toward fewer normative corrections.

Besides the unique character of each type, the findings strengthen what is already known from earlier research on the written and spoken language, thereby allowing us to determine the characteristics of each medium.

The spoken texts examined are characterized by many repetitions, both explicit repetitions of syllables, words, phrases, and whole sentences and implicit repetitions, such as the repetition of pronouns incorporated in the verb or the use of various kinds of apposition. Moreover, false starts and self-correction are commonplace in spoken language. The speakers use discourse markers heavily, many of which are specifically for spoken language, such as exclamations, vocatives, various kinds of parentheticals, text organization markers, turn-taking markers, various kinds of validity markers, and so on. Sometimes, spoken language is characterized by various levels of expressiveness.

Many of these features are often changed or completely deleted in the written version. The formal written language, in the genres examined, complies with the normative rules of The Academy of the Hebrew Language, and, therefore, many grammatical corrections were found: person, gender, number, definiteness, correction of prepositions, conjunctions, subordinators, and also corrections of the verb's conjugation and tense. The formal written language that was examined attempts to be not only more "correct", but also of a higher register. For this reason, some words and expressions either unmarked for register or marked for low register are substituted. Similarly, expressions that are not exactly appropriate for the context in which they were mentioned are exchanged for their counterparts marked for higher register, which are, according to the editor, more appropriate for written text.

## **POLARITY IN LANGUAGE LEVELS IN LITERARY TEXTS**

**Yitzhak Shlesinger**

Literary texts concern life realities, which are reflected in a variety of language devices and usages. This study focuses on polar language varieties that characterize life and reality, as well as figures that function in this reality. With this aim in mind, I examine register differences that express high and low language level in texts. These are sociolinguistic

characteristics that can be judged by native speakers as grammatical and appropriate in various speech communities and different contexts and circumstances. Literary texts are usually depicted as “rich” and figurative with high language usage. Inasmuch as language constitutes target linguistic material for writers, this language is supposed to be characterized by a high degree of language awareness. However, an in-depth analysis of texts indicates that texts contain a number of discourse styles, such as fiction, dialogues, and reflection. Previous analyses show that the language of dialogue is of a lower register than that of fiction, because it mirrors spoken interaction.

The study corpus consists of a main text, comprising literary works by six modern authors (Shifra, Yavneh, Izakson, Harsegor, Gur, and Katz) and of a comparable text that is not marked as “high”, comprising journalistic writing by six *Ma'ariv* and *Yedi'ot Aharonot* writers (Kharif, London, Snitzer, Plutzker, Deutsch, and Alon).

The study hypothesis is that utterances characterized by “low” register will be found appropriate by children and less literate individuals, and utterances with “high” register will be typical of literate adults.

Stylistic features of high language usage, from the lexicon, morphology, and syntax, were selected and analyzed in the text. These included, for example, higher lexical alternatives and bound, versus analytic, expression.

Results are not entirely homogeneous; however, they point to a decline in language usage on the time line. Fiction, as predicted, has a higher level than dialogues. However, a number of writers tend to produce literary texts with a higher register, and others do not follow the fiction/dialogue dichotomy.

## SPOKEN AND POETIC LANGUAGE IN ISRAELI MODERN POETRY

**Tamar Sovran**

Biblical poetry has been the most influential model for Hebrew poetry through the ages. It has been an endless source for quotations and allusions

as well. Hebrew poets opened themselves to the forms and content of European classical and modern poetry only toward the twentieth century. However, the new contents and the new poetic forms were still written in an archaic, high-style, poetic Biblical Hebrew. The turn of the century was an era of renaissance for Hebrew literature. Headed by Bialik, new poets liberated themselves, gradually, from the “golden chains” of Biblical *Meliza* (high poetic style). Instead of quoting from psalms, the prophets, and other poetic Biblical texts, poets now took the liberty of juxtaposing words and expressions from all sources of Hebrew, including the more secular layers of Mishna and Talmud. This change was followed by another process, when Hebrew poets incorporated modes of expressions, taken directly from the newly revived spoken Hebrew, into their poems.

The present paper outlines the earlier stages of the first process (sections 1 to 3) and focuses on the second stage. The paper’s main claim is that poetry is based on a paradox: Expressed in language, poetry has to follow linguistic rules, morphological, and syntactic as well as semantic rules, but it strives at creating innovative modes of expression by deviating from and by breaking these very same rules. Poetry is sensitive to contemporary poetic norms, to linguistic norms, and to poetry’s cultural context. Based on these assumptions, the paper displays the various modes in which the revival of spoken Hebrew provided a new arena for poetic tensions and innovations, mainly by bridging old connotative Hebrew expressions with new, spoken, often vulgar, patterns of speech.

Critics have claimed that modern European poetry drastically distanced itself from the earlier romantic norms by following the new norms set by T. S. Eliot, who used “low” materials for similes and metaphors. Modern Hebrew poetry has adopted this mode of creating tension between the poetic and the mundane, by dramatic switching of registers and tones.

Section 4 of the paper displays the ways by which an alleged regular tone or argumentation can bare a poetic impact by small deviations from the expected line of conversation.

Section 5 concentrates on morphological innovations that echo processes prevailing in spoken language. The section also looks at ways of

breaking politeness codes, social norms, and sexual taboos. Tensions are also created by metaphors and expressions that combine distant materials from both the sublime and the earthly and vulgar. Thus, the young poet E. Nachmias Glass can tell about his love-hate relationship with the national poet Bialik, quoting from one of Bialik's famous poems and immediately calling the poet "you son of a bitch". Using an Arabic form of the outburst intensifies the emotional effect.

The following section (6) describes various modes of penetration of patterns of speech and speech acts to the textures of poems, for example, in a tiny poem by the late D. Pagis. The poem is a dialogue between a stranger and a boy: "Who is your father, Son? — The second bouquet to the left." The context of cemeteries is evoked in the preceding short poem. Hebrew poetry follows modern trends in world literature, basically of disrespecting traditional conventions and breaking taboos. However, it has unique characteristics, because spoken Hebrew is very young and has only recently become a source for creating linguistic clashes with the rich echoes of traditional written language. The paper as a whole shows how poetry reacts to the spoken language and how poetry uses spoken language as a powerful poetic tool.

## **FROM TRANSCRIPTION OF SPOKEN TEXT TO ITS REPRESENTATION ON A GRID SET**

### **II-II Yatziv**

The aim of this paper is to present a new method for analyzing spoken texts. In addition to defining the principles and advantages of such a method, the paper includes specific examples of analysis of spoken texts using this method. I employed the process of analysis of the syntax and the structure of spoken texts, termed "analyse en grille" in French, in my corpus-based research, being part of my PhD dissertation, *Méthodologies pour la description de quelques phénomènes syntaxiques de langue parlée : application à l'hébreu moderne* (2002).

A few basic guidelines concerning the type of transcription and its objectives had to be convened before texts could be transcribed. Subsequently, the recorded texts were transcribed. The research, which deals with lexical and syntactic characteristics of spoken language, is based on the study of long monologues. This explains my choice of an orthographic transcription of the texts rather than a phonetic one. The words are transcribed as they appear in the dictionary, and the transcription includes some non-literal expressions, also known as discourse markers like *em*, *e* and *tfu tfu*. In addition, the use of punctuation marks has been purposely avoided, thus allowing postponement of the choice pertaining to units of analysis (sentence, speech act, paragraph) during transcription. This approach enables the researcher to reach the end of the syntactical analysis and to determine whether this phenomenon repeats itself later in the text and whether it is distinctive of the speaker or of the type of text.

Once the transcription is completed, the syntactical analysis begins, and only upon completion of the analysis is the visual presentation grid of the texts made possible. This type of grid suggests that most utterances we produce when speaking are complete utterances, meaning uninterrupted ones. It further supposes that between utterances within spoken texts, a discrete cohesion occurs, unnoticeable at first glance. Nevertheless, it can be detected using the grid set method.

The analysis is the assessment of the text as a dual axis structure — the horizontal (syntagmatic) axis and the vertical (paradigmatic) axis. By means of the grid set method, the linguistic material is spread on both axes. The syntagmatic axis is applied to the syntactic clause and its components (subject, predicate, and verbal complements). Each syntactic clause corresponds to one line of grid. The paradigmatic axis, on the other hand, is applied either to syntactic elements possessing an identical grammatical function or to hesitations, repetitions, or corrections, characteristic of spoken texts. These components are absent from the methods used for analyzing written texts. In contrast, the visual presentation method takes into account the complete syntactic clauses that structure long texts, as well as considering their mode of change during speech.

## POPULATION SAMPLING FOR THE ESTABLISHMENT OF A REPRESENTATIVE CORPUS

Giora Rahav

We are trying to find a way to represent the spoken language by means of a limited set of records of the language. One of our underlying assumptions is that language reflects both the speakers and the speech situations. We, therefore, try to sample the spoken language by means of a representative sample of speakers. Presumably, once we have such speakers, with samples of their speech lasting for a 24-hour period (spread throughout the week), we will have a proper sample of both speakers and situations. If we sample, from these situations, any 1-hour period (again, a representative sample), we will still have a representative sample. The major issue, then, is establishing a sample of speakers.

There are two major approaches to population sampling. The first approach is non-statistical or directed. There are many ways to do this, for instance, deciding that the sample should have X persons of Y origin in the M-N age group. Alternatively, one may start with any wide range of speakers, asking each speaker to nominate several other potential participants.

Statistical sampling is more difficult, but it has several advantages. These procedures involve an element of randomness into the selection process. This randomness makes certain that every speaker has a similar chance of inclusion in the sample. Some systematic deviations from this principle are allowed and may be used to ascertain the inclusion of the most significant social strata. One major advantage of these techniques is that the size of the sampling error (the deviation of the sample from perfect reflection of society) can be assessed.

However, the statistical approach forces us to define the population we sample very well and the way it will be accessed. One way of getting access to potential informants is to sample the records of the Ministry of the Interior and try to locate individuals. Alternatively, we may sample

addresses (cities, streets, buildings, and apartments) and then sample a suitable person within each address.

In addition, one must take into consideration that potential informants will have to be persuaded to cooperate. This necessarily means a high rate of sample attrition — persons dropping out of the sample. To make certain that the sample is representative, this attrition may significantly increase the costs of the survey. Thus, certain screening procedures will have to be devised to reduce this problem.

**THE PREPARATORY MODEL OF  
*THE CORPUS OF SPOKEN ISRAELI HEBREW (COSIH)***

**Benjamin Hary and Shlomo Izre'el**

This paper describes the initial design of *The Corpus of Spoken Israeli Hebrew (CoSIH)*. *CoSIH* will attempt to include a representation of most varieties of spoken Hebrew as it is used in Israel today. *CoSIH* is designed to consist of two complementary corpora: a main corpus and a supplementary corpus. The main corpus will form the bulk of *CoSIH* and will comprise about 95% of the entire collection. The great part of the main corpus (90% of the entire corpus) will be sampled statistically. For analytical purposes, it will use a conceptual tool in the form of a multidimensional matrix, combining demographic and contextual tiers. The combined demographic and contextual design will be capable of showing the distribution of speech types in various subgroups of the population. The remaining 5% of the main corpus will consist of targeted, demographically sampled texts, which will add to the randomly sampled data from special groups and which could not be included in the random sample. The supplementary corpus will include about 5% of the collected data and will add to the demographically based corpus a contextually designed collection on bases that we hypothesize are significant to the linguistic situation in Israel. This design is culturally dependent to suit the special structure of the Israeli Hebrew speech community and, thus,



includes both native and non-native speakers of Hebrew. Nonetheless, the principles governing this design are such that they would service the study of many other speech communities as well. The design itself may be employed for other corpora with only slight modifications. Further accounts of the *CoSIH* model can be found on the internet at <<http://spinoza.tau.ac.il/humanities/semitic/cosih.html>> and in Shlomo Izre'el, Benjamin Hary, and Giora Rahav, Designing CoSIH: The Corpus of Spoken Israeli Hebrew, *International Journal of Corpus Linguistics* 6/2: 171–197.

### **METHODOLOGICAL NOTES ON CREATING THE CORPUS OF SPOKEN ISRAELI HEBREW (COSIH)**

**Regina E. Werum**

This essay discusses the methodological considerations shaping the study design and data collection phases of *The Corpus of Spoken Israeli Hebrew (CoSIH)*. After briefly reviewing the scope of the project, the chapter focuses on sampling and analytical issues.

Substantively speaking, this project is unique. Unlike most existing corpora, the project focuses on the use of *spoken* Hebrew, thus echoing Fishman's call to shift from describing "model language" to mapping how we use language on a daily basis (Fishman, Introduction: The Sociology of Language in Israel, *International Journal of the Sociology of Language* 1, 1974, 9–13). Moreover, in most existing corpora, *words* and *phrases* form the primary phenomenon of interest, because such data sets are usually designed by/for grammarians and etymologists. In contrast, *CoSIH* focuses primarily on the *social context* in which Hebrew is spoken.

Methodologically speaking, *CoSIH* is also remarkable for its innovative, yet rigorous, data collection efforts. It combines the logic of both qualitative and quantitative research and facilitates pursuing different research goals simultaneously. *CoSIH* data will be highly useful to

researchers from the humanities, as well as the social sciences, because it allows us to (a) extract *descriptive* information, for example, about the ethnic or regional differences in speech; (b) answer *interpretive* questions about contextually significant phenomena; and (c) examine *causal* relationships regarding issues of social stratification.

*CoSIH*'s cross-disciplinary appeal is closely tied to its complex research design. Based on the premise that situational context shapes the use of speech, *CoSIH* data collection efforts revolve around three theoretically important axes. These "contextual variables" mark the relationship between the speakers, discourse structure (especially whether there is a power differential among speakers), and discourse topic. In addition to containing a variety of monologues, dialogues, and conversations, the resulting data set also contains media accounts, political speeches, and speech situations in other similar environments. All recorded speech segments lend themselves to deductive, as well as inductive, analyses.

Among linguistic researchers, little consensus exists regarding corpus construction. *CoSIH* differs from most corpora, because it does not employ convenience sampling and relies on an unprecedented sample size of  $n=1,000$  speech segments ("cells"). *CoSIH*'s innovativeness further extends to combination of sampling procedures, that is, systematic *random sampling* and *purposive sampling*. Designers faced the challenge of constructing a sample large and diverse enough to obtain a representative cross-section of spoken Hebrew, while also maintaining a thematic focus and ensuring that the data could yield parsimonious coding schemes. They decided to divide the sample into two sub-corpora, each of which illustrates a different type of sampling. The main corpus comprises 95% of the entire corpus: 950 linguistic "cells", each containing approximately 5,000 words of text. It relies on systematic random sampling, in which speakers are selected based on their place of residence. The subsidiary goal is to obtain a sample that is diverse in terms of the speakers' ethnic background, age, and educational level. In contrast, the supplemental corpus will consist of 50 additional instances of spoken Hebrew. Here, speech segments are not selected based on demographic criteria. Instead,

cases are selected based on the premise that some forms of speech disproportionately influence overall speech patterns, despite relatively few people actively disseminating such speech (popular media, parliament, and the court system).

The remainder of this essay discusses specific measurement issues, focusing on the challenges inherent in defining and operationalizing “ethnicity” and “power”, both of which form core social scientific concepts and illustrate the importance of social context in shaping speech patterns. The conclusion links the purpose of *CoSIH* to existing sociolinguistics and social psychology research that illustrates how related power differentials (e.g., by class or gender) affect language use and speech patterns.

To summarize, *CoSIH* has much to offer to social scientists. It treats speakers both as products of their environment and as producers of social reality. It allows explicit comparisons among various population groups. Finally, its sample size and complexity make it ideal for sociologists interested in language as a form of cultural capital, — a “currency” whose value depends on, for example, the personal background of the speaker and the situation in which it is used.