5 Analysing GERT data

First, the GERT corpus will be analyzed to determine core categories for linguistic variation in MH. An in-depth analysis of these categories follows by contextualizing them with interview data.

After the second fieldwork stage, I had twenty one GERT templates filled out by different participants as described in 4.2.6.3. While comparing them, the most obvious difference I could spot was the differing number of entries and varied distribution on the templates. It seemed that the participants had either filled in many entries or very few. The participants who only filled in up to four entries made more extreme statements by putting them in the opposite segments or by distributing them over all four segments of the template – they did not group multiple entries within one segment and left others empty. In all probability, two different strategies of categorization were used by the participants during GERT. A possible typification of the participants along these lines will be discussed in 5.6.

Fig. 5.1: Original GERT tokens as word cloud

https://doi.org/10.1515/9783111390628-005

Besides this initial observation, there was nothing that caught my attention right away when I was looking for emerging patterns in the templates. All original tokens from all GERT templates are visualized as word cloud in Fig. 5.1. The visualization is weighted: more frequently mentioned tokens appear in a relatively larger font. The aim of the following analysis is to make sense of this word cloud. To get a more encompassing and orderly perspective on the data, I transferred the information from all templates into a spreadsheet: The resulting GERT corpus is my primary data source for the following analysis. In the following sections, I will use the data of two participants as an example to explain the general analytic processes. I selected participants m69f4l2 and s41m3l1 because I found them to represent two different strategies of categorization. Most participants completed GERT in a very similar fashion to either one of these participants. The data of the other nineteen participants will be analyzed in conjunction and will only occasionally be discussed individually where a contextualization of the data is needed.

There are two possible approaches to analyze the GERT corpus: a qualitatively oriented analysis of categories, semantic domains and conceptual levels and a quantitative analysis of the number of entries, average values and the summary of these values in diagrams. Both approaches build on each other: the qualitative approach is the basis for defining and understanding quantifiable data (Mayring 2015: 20–22). In GTM, qualitative methods precede any quantitative analysis because they determine the research questions, the categories for the investigation and are used recursively to explain quantifiable data. Some form of quantitative analysis is needed for any generalization – it is important to show that similar patterns are recurrent in the data (cf. Mayring 2015: 53 and Nassehi 2019: 16). The results of the quantitative analysis need to be seen in the light of the original research questions and interpreted accordingly. I argue that some careful statements about the prominence of certain categories in the speakers' representation of linguistic variation can be deduced by counting and comparing participants' categories.

In the next sections, preliminary methodological thoughts for the analysis will be laid out before a detailed account of the GERT corpus will be given in 5.4. 5.4.3 provides an overview of participants' most frequently used terms. Participants' ratings of the terms are analyzed in 5.5.

5.1 Hermeneutic considerations for analysis

Taking a closer look at s41m3l1's template (see 4.10) will give us an idea which different types of categories participants used for the task and how they can be translated, interpreted and summarized. It will be discussed that these analytic processes are not forcibly linear and that their outcomes are more ambiguous than it may seem.

More evidently than for the interpretation and the summary of the data, the process of translation requires an analysis of linguistic structure. In MH – just as in other Semitic languages – there is not necessarily a structural difference between nouns and adjectives. *Maskil* functions both as head and as modifier of a noun phrase (NP): it can potentially be translated as 'educated person' and as 'educated.' In the context of GERT, s41m3l1's entry *maskilim*, which contains the masculine plural ending *-im*, is probably meant as 'educated persons' because it stands alone on the template, which makes its use as a modifier seem unlikely.

To determine the meaning of the entry for s41m3l1, in the context of its use during GERT, *maskilim* can further be analyzed as metonomy, which is "one of the basic characteristics of cognition" (Lakoff 1987: 77). By using the expression *maskilim* when referring to a group of people, one of their many characteristics is singled out and used to categorize them as one group – those with a high level of education. Lakoff generally describes metonomy as follows:

> It is extremely common for people to take one well-understood or easy-to-perceive aspect of something and use it to stand either for the thing as a whole or for some other aspect or part of it. (Lakoff 1987: 77)

The metonymic process of categorization which is documented as s41m3l1's performance during GERT depends on further premises which can be understood as belonging to an idealized cognitive model (ICM), in Lakoff's (1987) words. An ICM of 'education' can be modeled tentatively with these premises:

> (1) people can set themselves off from each other by the internalization of knowledge.[1] (2) Formal education, which is institutionalized as schools and universities, serves the systematic transmission of knowledge.

It will be assumed that this or a similar ICM determines the structure of categories such as *maskilim* 'educated people.' According to (2), it is likely that people who were exposed to more formal education than others have successfully internalized more knowledge which, based on (1), sets them apart from the others as 'educated people.'

Is that what s41m3l1 meant? When thinking deeply about 'education' it turns out that it is a very vague concept: to what extent are religious, moral, scientific and more practical aspects of knowledge – such as speaking according to linguistic norms – included in the concept and how are these aspects handled to determine if a person is 'educated?' The above quotation posits that metonomies function on the basis

1 This concept is expressed in the Talmudic verse: אינו דומה שונה פרקו מאה פעמים לשונה פרקו מאה ואחד 'one who reviews his studies one hundred times is not comparable to one who reviews his studies one hundred and one times.' (Chagigah 9b; citet with translation from www.sefaria.org)

of "well-understood or easy-to-perceive aspect[s]." Does this mean that a person's 'level of education' is easily perceivable for s41m3l1 – and if so, on what basis? In fact, because he used this concept during GERT, which asks participants to name categories for the distinction of HSs, s41m3l1 argues that his everyday understanding of 'education' enables him to categorize people based on their speech.

By definition, ICMs do not forcibly "fit one's understanding of the world" (Lakoff 1987: 70). Someone who was exposed to many years of formal education does not necessarily fit the category 'educated people:' the educational institution he frequented may not have qualified as 'good school' or he may have displayed a very passive attitude during his school years.

> Still we can apply the concept with some degree of accuracy to situations where the background conditions don't quite mesh with our knowledge. And the worse the fit between the background conditions of the ICM and our knowledge, the less appropriate it is for us to apply the concept. The result is a gradience – a simple kind of prototype effect. (Lakoff 1987: 71)

This means that, regardless of the appropriateness of its underlying ICM, the concept 'education' can still be used in every-day contexts. Its prototypical structure makes it operational for every-day contexts of categorization.

Insights into what is meant by the participants with 'educated,' for example, can be gained by analyzing interview data recorded during GERT. However, the perceptual basis for the participants' categorization – that is to say, what exactly makes a HS sound educated – can only be determined by perception experiments (see 2.1.4), which are not part of this study. For analyzing the GERT corpus I assume that most of the participants' entries can be summarized under a semantic domain. Thereby, participants' most used concepts for categorization during GERT will be singled out. The entry *maskilim* 'educated people' can be labeled as belonging to the semantic domain 'education.' Using a single semantic domain to label the loanword *giḳim* 'geeks,' which is another entry that was used by s41m3l1, is more complicated. It is used in MH, as well as in its source language English, to refer to people who are characterized by their enthusiasm for fantasy novels, role play games, and the like.

M69f4l2's entry *aḳadema'im* 'academics' can also be understood from at least two different conceptual perspectives: in terms of 'education,' as everybody who was exposed to some degree of university education or as indicating an 'occupation' and thus referring to everybody who is currently employed in academia as scientific staff. The two concepts are overlapping, but not identical, because not everybody with university education is currently employed as scientific staff. To clarify the exact meaning of *aḳadema'im*, as intended by m69f4l2, I should have asked her which of the two concepts she was primarily referring to. Without this information, I had to

make an analytic choice and interpreted it as belonging to the semantic domain of 'education'.

As Kuckartz (2016: 19) points out, these kind of interpretations happen subconsciously during a regular conversation and problems surface only when misinterpretations and subsequent misunderstandings become obvious. The lengthy discussion about inter-subjective differences in the interpretation of meaning did not yet include the important aspect of intercultural comparability of concepts: the similarity of concepts cannot be assumed categorically for different cultural contexts. With these reflections, I want to point out that my way of analysis is by no means definite or the only correct one. It can at best be an appropriate interpretation because there is no methodology to guarantee a correct interpretation (Kuckartz 2016: 20).

Necessary interpretations have consequences for the following analysis. I could have chosen 'occupation' as the corresponding semantic domain for *aḳadema'im* and consequently this domain would have received a more prominent position in the quantitative comparison of the participants' mentions of certain categories. It has to be kept in mind that the semantic domains 'education' and 'occupation' are conceptually related and that both were used frequently by HSs for their categorization during GERT, as will be shown in the quantitative analysis in Section 5.4.3. To conclude this hermeneutical excursion, it is noteworthy that a careful statement such as the one in the preceding sentence does not pose any methodological problems – even when considering the underlying possibility of misinterpretations.

5.2 Determining the relevance of categories

This study aims to shed light on conventions about the categorization of HSs and variation in MH. For the following analysis, it is assumed that the recurrent use of a category during GERT by several participants increases its likelihood to be relevant for the categorization of HSs and variation in MH, at large. This hypothesis is based on the cognitive principle of 'conventionalization,' as outlined by Schmid who defines 'convention' as

> a mutually known regularity of behaviour which the members of a community conform to because they mutually expect each other to conform to it. (Schmid 2020: 88)

He understands 'conventionalization' as subject to the cooperating and "partly complementary" processes of 'usualization' and 'diffusion' which he defines as follows, in the context of utterance types:

> The process of diffusion affects the numbers of speakers and sizes and structures of communities which partake in a cotext-dependent and context-dependent convention. It makes utterance

> types more or less conventional in the sense that more or fewer speakers or groups of speakers conform to a convention in a larger or smaller number of cotexts and contexts. (Schmid 2020: 93)

Based on a conceptual transfer of this notion of diffusion to the context of GERT, the recurrent use of a category by several participants is understood to hint at its higher degree of conventionalization in relation to categories which were used by fewer participants or not at all. Before the entries from the GERT corpus can be related systematically to concepts, more thoughts about their conceptual structure and the implications for this classification are due.

5.3 The nature of categories and levels of categorization

A brief look into the data reveals that participants behaved inconsistently in their use of categories and varied frequently in their degree of specification. Instead of mentioning multiple categories from the same semantic domain with a similar degree of specificity, participants used several semantically remote categories which may also differ in their degree of specificity. After mentioning *maskilim* 'educated people,' s41m3l1 did not proceed within the same semantic domain 'education' by indicating the opposite category on the same conceptual level *lo' maskilim* 'uneducated people.' Instead, he went on to conceptually mixed and more specific categories such as *ḥevra 'aravia maskila* 'educated Arab society' and its opposite *kfarim, ḥevra 'aravia lo' maskila* 'villages, uneducated Arab society.' The analyst may be puzzled by the nature of s41m3l1's categorization; however inconsistencies are known to be a natural and characteristic component of human categorization.

> There is in fact some evidence that natural conceptual hierarchies are fairly messy and not organized in a particularly consistent manner. [...] Furthermore, conceptual hierarchies do not even seem to be stable: there is evidence from attribute-listing experiments that categories may move from the subordinate to the basic level when they gain in cultural importance [...] Words such as (motor)car or (air)plane, for instance, which started out as subordinates in the field of vehicles, have since clearly acquired basic-level status. (Schmid 2007: 126)

Figure 5.2 shows some of s41m3l1's original entries, depicted as blue ellipses, and shall serve to illustrate their syncretic nature. The entry *kfarim, ḥevra 'aravia lo' maskila* 'villages, uneducated Arab society' includes references to the designated group's geographical location, their ethnic origin and their level of education. It can be linked to at least three semantic domains on a higher conceptual level, which are depicted in blue rectangles.

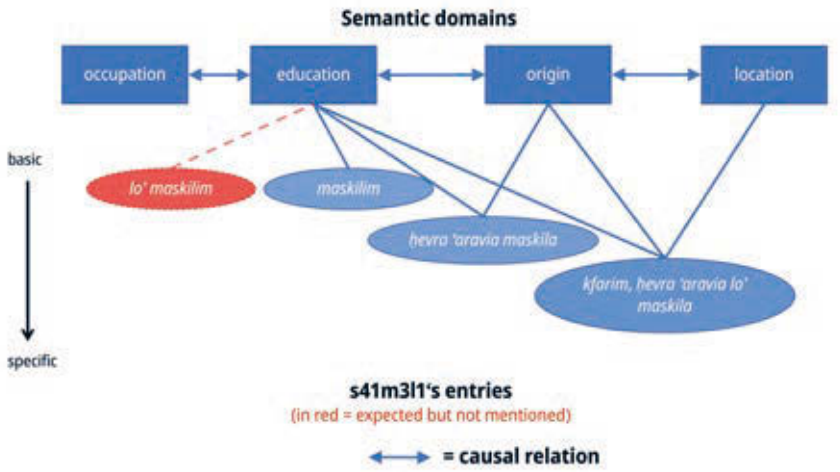

Fig. 5.2: S41m3l1's entries and their possible classification

The semantic domains themselves are not clearly separable, but semantically related (indicated by a blue double arrow). By their juxtaposition s41m3l1 expressed a correlation between the category 'uneducated Arab society' and the category 'villages.' While 'villages' can be understood as referring to geographical locations, 'being Arab' is not only determined by someone's current place of residence – which can be an Arab village. An additional semantic domain of 'origin' will be used in the analysis which can be understood as family descent, from a historical perspective: the place of residence of past generations is determining someone's origin. In Israel, there is also a correlation between educational aspects, 'origin' and 'location' because Arabs usually attend Arab schools. The most obvious relation extends between 'education' and 'occupation' because people's occupational choices are often determined by educational criteria. The syncretic character of many entries makes it hard to classify them under a single semantic domain. To get a better understanding of participants' categorization processes documented with GERT, it is helpful to reassess the entries with Rosch's definition of 'category:'

> [C]ategories tend to become defined in terms of prototypes or prototypical instances that contain the attributes most representative of items inside and least representative of items outside the category [...] By category is meant a number of objects that are considered equivalent. Categories are generally designated by names (e.g., dog, animal) [...] (Rosch 1978: 30)

Schmid further elaborates on the potentially syncretic nature of natural categories:

> If the logical principle of class inclusion is declared invalid – at least for natural conceptual hierarchies – as a determinant of category status at the vertical level, this has consequences on the horizontal level as well: categories at the same level of categorization need not always be mutually exclusive. (Schmid 2007: 127)

This understanding of 'category' is substantiated by the characteristics of the participants' categories which have been analyzed, so far. From a general perspective, Kristiansen (2008: 72) argues for the prototypical nature of categories such as "[a]ccents, dialects and social stereotypes," of which most GERT entries are instances. Krefeld & Pustka also apply the notion of prototypicality to varieties:

> [V]arieties are located in the minds of the speakers, in that they are characterized by a prototype structure, as all other semantic concepts from our everyday life are. These cognitive concepts are not only expressed in language use but language use combined with para- and non-linguistic factors, co-construct situations and identities. (Krefeld & Pustka Forthcoming)

The organization of categories along prototypes implies that category membership can at best be approximated, but not determined definitely. Therefore, alternative analytical possibilities need to be accounted for, but should not pose a methodological obstacle.

In the above quotation from Schmid (2007: 127), the notion of different conceptual levels is taken up, which is seen as further characteristic inherent to categorization processes (Rosch 1978: 30). This aspect is reflected in s41m3l1's original entries by their differing degree of specificity which is visualized in Fig. 5.2 with a black arrow on the left margin. S41m3l1's entries contain categories which belong to several conceptual levels: among the entries are basic categories such as *maskilim* which can comprise entries that refer to more specific categories, just as *ḥevra 'aravia maskila* 'educated Arab society.' The category *kfarim, ḥevra 'aravia lo' maskila* 'villages, uneducated Arab society' is even more specific because it refers to certain villages and highlights two of their inhabitants' characteristics: 'being Arab' and 'being uneducated.' In comparison, *maskilim* 'educated people' is less specific and comprises a lot more people.

Because the original entries belong to different conceptual levels, it is problematic to compare them directly: they have to be brought to a similar conceptual level before the next analytical step. Following Rosch's (1978: 30) notion of a 'basic level of categorization,' it will be assumed that conventionalized categories – the categories which I was trying to elicit with GERT, are close to a basic level, which makes this level a good point of departure for further comparisons. Harder describes the relevance of this basic level for the study of representations:

> The basic level is thus a fairly solid new point of departure for understanding the kind of mental representations that real people construct: down-to-earth, no more precise than required for everyday life, capable of accommodating a broad spectrum of different cases, associated with practical as well as conceptual skills. In short, they reflect both properties of conceptualization as a human skill and properties of its basis in experience. Basic level concepts are shaped by an economy factor: they end up at a level of generalization and specificity that balance out costs and benefits of cognitive efforts. In relation to human experience, they also reflect the patterns of co-occurrence in the phenomena that constitute the input to conceptualization [...] (Harder 2010: 20)

RQ1 and RQ2, which ask about HSs' mental representations of linguistic variation and social groups, can be tackled by analyzing a basic level of categorization in these respects. A method to determine a basic level for further classification of the GERT corpus will be developed in 5.4.1 and 5.4.2.

Reflecting on the reasons for the conceptual inconsistency of s41m3l1's entries, one can argue with Rosch's (1978: 28) second principle of categorization which asserts that categories are dependent on the "perceived world structure:"

> What attributes *will* be perceived given the ability to perceive them is undoubtedly determined by many factors having to do with the functional needs of the knower interacting with the physical and social environment. One influence on how attributes will be defined by humans is clearly the category system already existent in the culture at a given time. (Rosch 1978: 29)

The importance of language in this respect has been discussed above with Berger & Luckmann's (1967: 22) argument that "language marks the co-ordinates of my life in society and fills that life with meaningful objects." One can easily see that the list of s41m3l1's entries (see Fig. 5.3) contains several culturally determined concepts, such as 'Haredim' which are only meaningful in the context of Israel. Regardless of the relation of the underlying concepts to actual events or experiences, the existence of conventionalized lexical items such as *ḥaredim*, *ashkenazim* and *mizraḥim* increases the likelihood that these terms, instead of others which may be more appropriate, are used in categorization processes.

Various other aspects which are hard to determine exactly are likely to influence s41m3l1's choice of words during GERT. Perhaps he did not want to use the plain term for 'uneducated people' because he conceded to conventions of political correctness. Maybe he avoided the term 'uneducated' because he thinks of himself as someone who occupies a privileged position as an academic and does not want to look down on less educated people. It is probable that the more specific entry 'villages, uneducated Arab society' let him feel more at ease because the problematic notion 'uneducated' is embedded within other attributes and it is not his own society but Arab society which may enable him to judge with a sort of detached attitude.

S41m3l1's choice of words is also influenced by the way he constructs his identity in relation to certain social groups of which he believes himself to be a member – or an outsider. He was the only participant who mentioned *giḳim* 'geeks' because he self-identifies as a member of this group. The entry *giḳim* can be found on his template (see Fig. 4.10) in the upper right segment in an ellipse, right next to the point where the participant located himself with an X and the entry *'ani* 'I.' The third entry in this segment above the ellipse and right next to his self-referral is *maskilim* 'educated' – another category which he chose to describe himself.

From these observations and the theoretic discussion in 2.1.2.2, it can be seen that participants did not produce their entries during GERT following a strict taxonomy, but according to how they want to convey their own identity in relation to their social and cultural environment.

5.4 Summary and explication of the data

To create a spreadsheet for the analysis of the GERT data, I transferred the information from the templates and added columns for translations, summaries and meta data which allow for the reorganization of the data. The different methods of summarizing and structuring the data obtained with GERT are the basis for their comparison and their quantitative analysis laid out in 5.4.3 and 5.5.

5.4.1 Defining units and types of data

As starting point, I collected all the data from all the completed GERT templates and transferred them into the first column of the spreadsheet in Hebrew orthography with their original spelling. In the following, I am going to render the original entries in the text in the form of a transliteration. For this transfer process, I worked with one template at a time and started by scanning the right segment on top from its upper right corner to its bottom left corner before proceeding counterclockwise to the next segment. Because the information on the templates appears in its majority as easily distinguishable entries (see Fig. 4.10), it felt intuitively appropriate to choose these entries as the basic unit for analysis. I transferred every entry into a separate row in the first column of the spreadsheet. In a separate column, I entered the corresponding participant's siglum.

In cases of doubt regarding how to segment information from the template into units, I applied three criteria for their distinction: spatial, semantic and procedural. The most obvious criteria is the spatial distribution of information on the template. Single tokens which are spatially separated from each other qualify as

entries. Additionally, participants sometimes encircled their entries, so they can easily be distinguished as units. However, if two or more tokens are close, it has to be decided if they belong to a single or to several entries. In this case, I applied semantic criteria to decide whether the second token is more likely to be an attribution belonging to the preceding one or an entry on its own behalf. For example, it is straightforward that s41m3l1 referred to '(female) social workers' with the two tokens *o'vdot sotsia'liot* because *o'vdot sotsia'liot*, which is a loan translation from English, constitutes a lexical item in Hebrew as well in English (Rosenthal 2009: 676). *Sotsia'l-i-ot* 'social-ADJ-F.P' does not qualify as a separate unit. During the analysis, I tried to recapitulate how the participant had filled out the template in my presence. Therefore, I listened to the recordings of the task. For example, I asked myself if the participant wrote down the tokens in close succession or with a pause in between them and whether he talked about them in conjunction or on their own terms. I am subsuming these aspects as procedural criteria. Listening to the recordings of the task also helped me to decipher responses when I had difficulties reading the handwriting or didn't remember what a participant possibly wanted to indicate. For the later analysis it was an advantage that I had reviewed every template on the day of its completion and made notes on the template, which can be seen in Fig. 4.10, whenever I felt an explication was needed. I will keep on using the term "entry" to refer to the now defined basic units for analysis. Not all entries are categories because entries can contain several categories or refer to specific places or persons. Therefore, I will not use these notions interchangeably.

Among all 190 entries that I transferred from the templates, three different types of entries can be defined according to formal and semantic criteria. An overview of these types, their qualities and the number of entries for each type is given in Table 5.1. Type A contains all 89 single token entries, except single token entries referring

Tab. 5.1: Types of entries according to formal and semantic criteria

Types	Tokens	Sem. criteria	n	Ratio in %	Recurrent entries
A	1	Lexical items without modifier	89	47	60
	2	Compound lexical items without modifier	17	9	6
B	≥ 2	Phrases including modifiers and/or clauses	65	34	2
C	≤ 3	References to specific persons and places in Israel	19	10	2

to a specific person or a place in Israel – those belong to type C – and 17 two token entries which are lexical items such as the above mentioned *o'vdot sotsia'liot*. All entries of type A, which account for 56% of the entirety, have in common that they

cannot easily be paraphrased in MH with a structurally more basic non-composite or a shorter wording.[2] In accordance with Coates' (2006: 371) definition, they can be described as common nouns which are used by the participants for semantic reference to a class – in our context a group of people – rather than for onymic reference to a specific entity, such as a specific person. In terms of construction grammar, all entries of type A are constructions because "their form or meaning is not strictly predictable from the properties of their component parts or from other constructions" (Goldberg 1995: 4). From this definition follows that type A entries cannot contain modifiers of any kind. These defining properties of type A characterize the terms which were used most frequently by the participants to refer to distinguishable groups of HSs.

Among the 106 entries of type A, there are 66 entries which appear more than once in the data. Minor orthographic differences, such as alternative and wrong spellings of a term which already appeared in the corpus as well as terms preceded by the definite article are included in this classification. The majority of recurrent entries appears in the corpus with identical orthography and more than twice: there are three terms which appear twice, six terms appear each three times, four terms appear four times and each one term appears five, six, seven and eight times. The number of appearances of each term equally indicates the number of different participants who used the term during GERT. While more than half (62%) of type A entries were used by at least two participants, only one entry of type B and one entry of type C appeared twice. The recurrent entries will be analyzed in detail in 5.4.3.

All 65 multi-token entries which are defined as type B contain between two to twelve tokens. By definition, they are structurally and semantically more complex than type A entries. For example, the type B entry *migzar 'aravi* 'Arab sector' is a two token composition which can be used to refer to the same concept as the type A entry *'aravim* 'Arabs.' The term *migzar 'aravi* is used in Israel to refer to the Arab population in a more sophisticated way than using the plain term *'aravim* – it has a specific pragmatic component which is not present in the plain term. Some type B entries like the above example can potentially be paraphrased with a more basic wording. Many other entries which include modifiers and clauses refer to a semantically specified concept which already appeared as a type A entry in a more basic form. The NP *ḥevra 'aravia maskila* 'educated Arab society' is essentially a semantic specification of the concept expressed by *'aravim* 'Arabs.' Its head *ḥevra 'aravia* 'Arab society' is a similarly elaborate term for 'Arabs,' just as the one discussed above. The dependent adjectival modifier *maskil-a* 'educated-F.S' functions as a semantic specifier which modifies the NP to refer only to the educated subgroup

2 All the ratios here and in the table are rounded half to even.

among Israel's Arab population. Their inherent complexity reduces the likelihood of type B entries being used by different participants. In fact, the only type B entry that appears twice is *'olim mi-'etiopia* 'immigrants from Ethiopia' – a concept which was referred to by most participants with the basic term *'etiopim* 'Ethiopians.' Type B entries account for slightly more than a third (34%) of all the entries. This indicates that participants preferred type A entries (56%) for their classification, which are by definition more basic in form and meaning.

Type C entries are semantically defined as "onymic" references to specific persons or places, following Coates (2006: 371). They contain 14 references to a person and 5 references to a place in Israel. The 19 type C entries consist of one to three tokens. Entries referring to persons are seven proper names of different Israeli celebrities, five self references to the participants realized as a pronoun or the participant's name and two references to the interviewer by his name (the only recurrent term of type C). Only five entries refer to the geographical space in Israel: three toponyms (*Tel Aviv, Ḳrayot, Daliat al-Karmel*) and two demonyms, which are geographical characterizations of people just as the entries *toshvei Yeruḥam* 'residents of Yeruḥam' and *toshvei Hertseliya* 'residents of Hertseliya.' Demonyms and toponyms can in our context be used as categorical terms for groups of HSs. Referring to a specific person can be interpreted as referring to the person's main characteristics, thereby enabling comparisons and generalizations. Nonetheless, type C entries should not be interpreted as a categories, per se. Because of their specific type of reference, they will be preserved in their original form in the next analytical step, which combines the translation of the entries to English and the bundling of semantically similar entries from different participants.

5.4.2 Translation and simplification of entries

It was necessary to translate the entries to English to make the analysis accessible for non-Hebrew speakers. At the same time, I wanted to find out which kind of entries were used most frequently by the participants. To this end, I had to bundle entries which refer to the same or to a very similar concept. I was aware that I could not be sure how exactly two concepts match which were referred to by different participants using similar terms – even if they had used identical wordings (see 5.1). However, summarizing elicited data is used in psycholinguistic studies as analytical method (see Lavan 2023 who used a comparable method for classifying how participants categorized faces and voices). Every translation involves some kind of interpretation. It was impossible to preserve the exact meaning of the entries in their translation and a verbatim translation of all entries would have obscured the analysis. Consequently, I decided to combine translation and bundling of the entries within a single analytic

step. This process was neither straightforward nor linear: I have been revising the data over and over for a period of several months and discussed my analytic decisions with colleagues. Sometimes, I introduced changes only to fall back on the prior solution because I was not convinced. My analytic decisions will be explained in the following.

The technique of *Zusammenfassung* 'summary' is defined by Mayring (2015: 67) as reducing data, while preserving their essential content which leads to the creation of a manageable corpus that reflects the original data. The guidelines set by Mayring (2015: 71–72) were helpful for the bundling of similar entries – except that I did not yet want to bring all the data to the same conceptual level at this early stage of the analysis. Instead, I oriented my choices towards an intermediate conceptual level, to which most of the entries seemed to conform. When it seemed more appropriate, I preserved different conceptual levels from the original data. For example, the entries *sṭudenṭim* 'students' and *profesorim* 'professors' are not on one level with m69f4l2's entry *aḳadema'im* 'academics.' Instead of bundling all three entries under the more comprising term *academics*, I chose to keep *students* and *professors* as separate entries. This way, the participants' classificatory strategies are still reflected to a high extent in the edited data. All terms resulting from this process will be referred to as "simplification" (of the original entry) and were added to the spreadsheet as column, next to the entries.

Type C entries were transliterated from Hebrew to Latin script, but not further summarized. According to semantic and formal criteria which were defined above, all other entries can either be described as basic (type A) or as complex (type B). Type A entries account for the majority and more than half of these entries occur several times in the corpus. Obviously, these recurrent terms are relevant for several participants. Based on their characteristics, it seems natural to think of type A entries as prototypical and belonging to a basic level of categorization (see 5.3).

Because I wanted to preserve the major characteristics of the data, I selected already existing type A entries as simplifications for semantically similar entries. If no similar entry was available, I chose new terms as simplification which matched the type A criteria. Generally, I preferred single-token terms as simplifications over multi-token terms. While most entries of type A didn't have to be paraphrased for their bundling, type B entries usually needed to be simplified for the sake of comparability. Some entries of type B were not simplified because there was no similar concept among all entries. I started with easily translatable basic terms such as *'aravim* 'Arabs.' At the same time, the process of bundling similar entries began: Some type A entries were paraphrased in English, instead of translating them directly, because there was a semantically similar entry among them which was more representative for the summary. For example, *'universiṭa* 'university' was paraphrased as "academics" on the basis of the entry *'aḳadema'im* because it is a

similar but more comprising term to describe people who possess some degree of university education – the concept to which the participant most likely referred with *'universiṭa* (see 5.1 for a discussion of the term *'aḳadema'im*).

Subsequently, the resulting simplifications served as paraphrases for semantically similar but structurally more complex entries. For example, I paraphrased the two entries *ba'alei haskala 'universiṭa'it* 'those with academic education' and *'anshei haskala 'universiṭa'it* 'people with university education' which were mentioned by different participants as "academics." For illustration, a detail of the spreadsheet at this stage of the analysis is shown in Fig. 5.3. If simplifying an entry implied omit-

Original Entry	speakers's sigla	Simplification	Specification
אקדמאים	m69f4l2	academics	
דתיים לאומים	m69f4l2	National Religious	
חרדים	m69f4l4	Haredim	
צפונים	m69f4l5	Tel Aviv	North
תברה ערביה משכילה	s41m3l1	Arabs	educated
כפרים, תברה ערביה לא משכילה	s41m3l1	Arabs	low education, villages
צבא	s41m3l1	army	
צבא קבע	s41m3l1	army	fixed job
עולים מאתיופיה	s41m3l1	Ethiopians	Olim
גיקים	s41m3l1	geeks	
חרדים	s41m3l1	Haredim	
מהגרים שהם משכילים	s41m3l1	immigrants	educated
עולים	s41m3l1	immigrants	
משכילים	s41m3l1	intellectuals	
אני	s41m3l1	selfreference	
עובדות סוציאליות	s41m3l1	social workers (female)	
מורות	s41m3l1	teachers	F
פועלים	s41m3l1	workers (blue collar)	

Fig. 5.3: Summarized and translated GERT data from m69f4l2 and s41m3l1

ting semantic components, I preserved this information on the spreadsheet in a separate column under the title "specification." For example, s41m3l1 referred twice to 'army', thus differentiating between plain *tsav'a* 'army' by which he meant 'conscripted soldiers and reservists' and *tsav'a ḳev'a* 'standing army' referring to 'soldiers working for the army on a fixed basis' – as he explained during the interview. Both entries were simplified to "army" and *tsav'a ḳev'a* got transferred as "fixed job" in the specification column.

Some cases were more tricky than the ones just mentioned. I decided to use "Russians" rather than "immigrants" as simplification for a30f3l2's entry *'olim ḥadashim mi-brit ha-mo'atsot*, which translates to '(new) immigrants from the Soviet Union.' This choice may seem illogical because the term 'Russians' usually denotes a nationality which once belonged to the Soviet Union, but does not comprise other

nationalities, such as Kazakhs, which were part of the Soviet Union, too. This analytic decision makes sense if we consider the context of the interview with a30f3l2 and the participants' biography.

At the time of the interview in January 2020, the Soviet Union had been dissolved for almost thirty years. The participant used the term *'olim ḥadashim* as head of the NP, which translates verbatim to 'new ascendants,' but is conventionalized as a lexical item, denoting immigration to Israel based on the "Law of Return.[3]" Usually, *'olim ḥadashim* is used to refer to immigrants who started the immigration process recently, whereas *'olim ṿatiḳim* 'senior immigrants' denotes immigrants who completed the process and have been living in Israel for years. In fact, a30f3l2 distinguished between new and senior immigrants during GERT by the use of *'olim l'o ḥadashim* 'not new immigrants' in opposition to *'olim ḥadashim*. What a30f3l2 meant most likely are immigrants from former Soviet countries such as Ukraine, Kazachstan and the like who arrived in Israel recently and after the dissolution of the Soviet Union. I argue that the seven participants who used the term *rusim* 'Russians' during GERT referred not only to Russians who were born in the state of Russia, but to descendants of countries which once belonged to the Soviet Union. They used the term *rusim* to refer to this group of people on the basis of their main distinguishable feature – speaking Russian. This argument will be elaborated in Section 6.5 where the core category *rusim* 'Russians' is contextualized with interview data.

A30f3l2 had personal reasons to choose a different wording, when referring to the same concept as the other participants who used the term *rusim*. She identified as belonging to this social group, as she clarified during the interview: she migrated to Israel from a former Soviet country with her parents as a child in the 1990s. Instead of adhering to the common practice of referring to all descendants from countries of the former Soviet Union as *rusim*, she used a geographically more precise wording. Her choice of words can be interpreted as an expression of personal affectedness by the politically sensitive topic. In analogy to the alternative Hebrew terms which denote the Arab population (cf. 5.4.1), there is also a pragmatic component to the use of alternative terms for *rusim* which can be perceived as blunt. Although historically imprecise, a30f3l2's entry is more encompassing than *rusim* when referring to the concept of immigrants from post-Soviet states.

Some participants produced composite entries which contain more than one category, such as h21f3l2's *druzim 'aravim tsafon* 'Arab Druze North' and a20f2l2's *'olim 'afriḳ'aim be-klali* 'African Olim in general.' I had to decide whether to simplify these entries as 'Arabs' or as 'Druze' and in the second case as 'Africans' or as 'immigrants.'

3 This law includes the term in its masculine singular form *'oleh* and states that "[e]very Jew has the right to come to this country as an oleh" (Israel Ministry of Foreign Affairs 1950).

Based on "the general cognitive principle that special cases take precedence over general cases," Lakoff (1987: 74) suggests "that in conflicts between modifiers and heads, the modifiers win out" for the interpretation of "complex concepts." In this respect, *druzim 'aravim tsafon* is an untypical construction because *'aravim* 'Arabs' is the modifier, although *druzim* 'Druze' is a more specific concept – usually Druze are categorized as 'Arabs' in Israel (see 3.1.4). In this case, I simplified as 'Druze' and in the second case as 'Africans' because it is the more specific category to which a20f2l2 most likely referred.

I used four main strategies for translation and simplification of the original entries. They are summarized in Table 5.2 using the examples which were explained above. Some type B entries were also translated directly when there was no similar

Tab. 5.2: Main strategies for translation and simplification

	Technique	Condition of use	Example
1	Transliteration	only type C	ירון לונדון → 'Yaron London'
2	Direct translation	primarily type A	*'aravim* → 'Arabs'
3	Simplified translation	type A, type B	*migzar 'aravi*; *druzim 'aravim tsafon* → 'Arabs'
4	Paraphrastic translation	type A	*universiṭa* → 'academics'

and more basic concept in the corpus. The degree of interpretation that is involved in simplification increases from technique 1 to 4. While a transliteration is a direct transfer, a paraphrastic translation involves theoretical assumptions which allow to think of *universiṭa* in terms of 'academics,' for example.

5.4.3 Recurrent categories

GERT was designed as a tool to systematically elicit categories that HSs use to represent variation in MH. The hypothesis was introduced above that the relevancy of these categories can be tested with the quantitative analysis of their independent mentions (see 5.2). In the context of their elicitation of a corpus of swearwords, with a total number of 56 participants, Vallery & Lemmens argue that

> if a word is given spontaneously by two speakers as a swear word, it is probable that it is conventionally considered as a swear word by at least some proportion of the population. On the contrary, if a word is given by only one out of 56 respondents, then there is a much higher probability that no one else (or an insignificant amount of people) considers it to be a swear word. (Vallery & Lemmens 2021: 92)

The same argument can be made about the general relevance of the categories which were elicited with GERT. In fact, the smaller sample size for GERT, with 21 participants, reinforces the likelihood that a category which was mentioned independently by two or more participants is relevant for the research population. A greater number of independent mentions of a category further increases its relevancy.

Therefore, I decided to focus only on the entries for which I found an identical or a semantically very similar entry from another participant in the GERT corpus. In other words, I am using the simplifications of the original entries for the bundling of semantically similar entries and the comparison of the number of their mentions and I will focus only on simplifications that are at least linked to two different participants. Furthermore, I did not want to include multiple occurrences of a simplification from just one participant. As can be seen in Fig. 5.3, s41m3l1 referred to 'army' with two different entries. I shaded these multiple occurrences from the same participant in red in the spreadsheet and counted them just once. All 21 foreign-induced entries which were written with the blue pen (see 4.2.6.3) are shaded in blue on the spreadsheet, too. These entries also need to be excluded from the present analysis to be able to argue in terms of relevancy. Further details about the entries which were so far discarded will be discussed separately in 5.4.4. These contain 46 different terms for which no semantically similar term could be found in the process of translation and simplification – they are discarded as single mentions. Because the single mentions were given by just one of the 21 participants, it is unlikely that they are commonly used by HSs for the categorization of variation in MH or for groups of speakers. Two participants referred to me (the interviewer) with an entry during GERT. However, these entries are not proper single mentions and they will be excluded from the analysis as well because they do not qualify as a relevant category. After excluding all irrelevant entries, 106 simplifications are left which will be treated in the following as categories because they were used by multiple participants to refer to groups of people. Table 5.4.3 contains a summary of these simplifications yielding 25 semantically different categories. The left column shows the number of mentions by different participants for each category, whereas the categories are displayed to the right, separated by semicolons.

A semantic interpretation of these recurrent categories reveals that the three concepts 'origin,' 'education' and 'religion' were most prominent for the participants' categorization during GERT. Based on the differing frequency with which these concepts were used, there seems to be a conceptual hierarchy. 'Origin' is the most frequently used concept: almost all of the eight categories which were mentioned by at least six participants are based on the concept 'origin;' only the categories 'Haredim' and 'academics' which were mentioned by seven and by six participants relate to the concepts 'religion' and 'education,' which seem to be secondary concepts. The categories 'intellectuals' and 'uneducated,' which were each used by four

Tab. 5.3: Recurrent categories during GERT

Mentions	Simplifications (separated by semicolon)
10	Arabs; Russians
8	Ethiopians
7	Ashkenazim; Haredim; immigrants
6	academics; Mizrahim
4	intellectuals; uneducated; Moroccans; teachers
3	Druze; politicians; television and radio hosts
2	army; Arsim; Kibbutsniks; lawyers; Moshavniks; national religious; seculars; settlers; workers (blue collar); Yemenites
106	recurrent categories in total

participants, also belong to the concept 'education.' Among the rest of the recurrent categories, which were mentioned by four participants or less, are several which refer to 'religion,' such as 'Druze,' 'national religious' and 'seculars.' The categories 'Moroccans' and 'Yemenites' can also be grasped in terms of 'origin,' while some other categories such as 'kibbutsniks' and 'settlers' are not as easily comprehensible. They potentially refer at least to geographic, religious and socioeconomic aspects because they can best be understood as describing a way of life.

When comparing these categories to the variables which are generally considered for sociolinguistic studies (see 4.1.1), it seems astounding that most of the variables do not even appear. From Barron & Schneider's (2009: 426) list, the variables "region," "gender" and "age" are completely absent in the recurrent categories. Only "ethnicity" if understood as 'origin' was frequently referred to besides the optional factors from the list "education and religion." The category 'workers (blue collar)' is the only clear reference to the remaining variable from the list "social class" and was mentioned by just two participants.

My hypothesis that 'region' and 'social class' might be less significant factors in Israel than elsewhere – at least in speakers' representations – is substantiated by the analysis so far. It is surprising that no geographic place was referred to more than once in GERT. Jerusalem, which was mentioned as having a few shibbolets during the interviews, was not mentioned at all. The relative prominence of 'religion' justifies the inclusion and the prioritization of this variable for the analysis. Even though I also expected 'origin' to play a major role for HSs' representations of linguistic variation, its total prominence over other concepts is astounding. Perhaps, GERT is less apt for the elicitation of categories along the variables 'age' and 'gender.' Despite the fact that only one entry in total referred to 'age,' the aspect was addressed more often in the interviews. Also the variable 'gender' was addressed in conjunction

with special contexts such as differences between male and female L2 speakers and especially differences between male and female religious HSs.

The meaning and the cognitive nature of the most prominent categories will be further analyzed in contextualization with interview data and participants' ratings during GERT in Chapter 6.

5.4.4 Foreign-induced entries and single mentions

This is a short account of the data that has so far been excluded from the analysis: foreign-induced entries and single mentions. As explained in 5.4.3, I marked 21 entries as foreign-induced. This means that I mentioned these terms prior to the participant during GERT. Foreign-induced entries occurred with seven different participants – a third of the sample size. Among the 21 foreign-induced entries, there are seven which refer to the army. These entries stem from four different participants, two of which referred to the army with multiple entries. The second most frequent category among the suggestions with three mentions from different participants is 'Haredim.' The rest consists of each two mentions by different participants of the categories 'kibbutsniks,' 'Druze' and 'Tel Aviv' and one mention each of the categories 'national religious,' 'politicians,' 'religious' (referred to twice by the same participant) and the entry "Christians in Nazareth."

There are two obvious explanations for the occurrence of these suggestions: the most frequent foreign-induced entries occurred with the categories 'army' and 'Haredim' – two topics which I found particularly interesting and was determined to investigate further. Certainly, I was disposed to bring up these topics if the participants did not mention them by themselves. The same explanation applies to the other categories to a lesser degree. I mentioned some of the remaining categories because they were somehow close to the participant or the interview situation. Sometimes I mentioned an example of a social group to explain what I expected of the participants during GERT. To this end, I tried to find examples which were familiar to the participants.

The second type of discarded entries are single mentions. One example for a single mention is the term *giḳim* 'geeks' which was discussed above (see 5.1). There are six more single mentions for which no similar concept could be found as simplification. At 46, the amount of single mentions is considerably high, which is partly due to 14 entries which refer to real persons. Five of them are self-referrals to the participants, just as the one from s41m3l1 which was also discussed above. Twice the participants also marked my (the interviewer's) position on the template which does not qualify as a proper single mention, but surely does not qualify as a meaningful recurrent category, either.

Among the single mentions are the entries *bnei no'ar tsa'irim* 'teenagers, youths,' which was the only entry that primarily refers to the concept 'age,' and the only proper mention of *yehudim datim* 'religious Jews.' I would have expected to get these notions more often in GERT, especially because the participants referred to 'age' during the interviews as a variable for variation in MH. Almost all the other single mentions belong to the semantic domains 'occupation' with 11 entries and 'origin' with seven entries. These are also the semantic domains which were most often referred to during GERT. All of these entries will again be included in the following analysis of the participants' ratings.

5.5 Ratings of 'status' and 'correct Hebrew'

To analyze participants' ratings, I indicated on the spreadsheet where the participant had placed the entry on the template relative to the printed axes. I used two separate columns next to the entries for their position in relation to the axis "correct Hebrew" and to the axis "social status." As indicated in 4.2.6.3, the task was not designed to elicit precisely quantifiable comparisons between different entries. Originally, I just indicated the entries' positions as being on the positive or the negative half of the axis or very close to the middle: the value "1" corresponded to a position on the positive half of the axis, "–1" to a position on the negative half and "0" to a position right on or very close to the other axis in the middle of the template.

After consulting with colleagues, I decided that a finer grained method for the analysis was possible because I had instructed the participants to think of the space in each segment as carrying meaning for the comparison of their entries. An illustration of the scale with five distinct values applied on s41m3l1's template can be seen in Fig. 5.4. In practice, I marked the values in red on a transparent foil which I put on top of the templates to determine the values of each entry in relation to the scale. I decided to use the scale with the values "–2, –1, 0, 1, 2" because I could determine clear spacial distinctions along these values on most templates, as can be seen in Fig. 5.4 – but a finer distinction was not deducible in most cases. For example, s41m3l1's entry חברה ערביה משכילה (*ḥevra 'aravia maskila* 'educated Arab society'), which is indicated by a blue arrow in Fig. 5.4, received the values "1" and "0" because it is located in the middle of the 'correct Hebrew' axis and closely to the value "1" on the 'social status' axis.

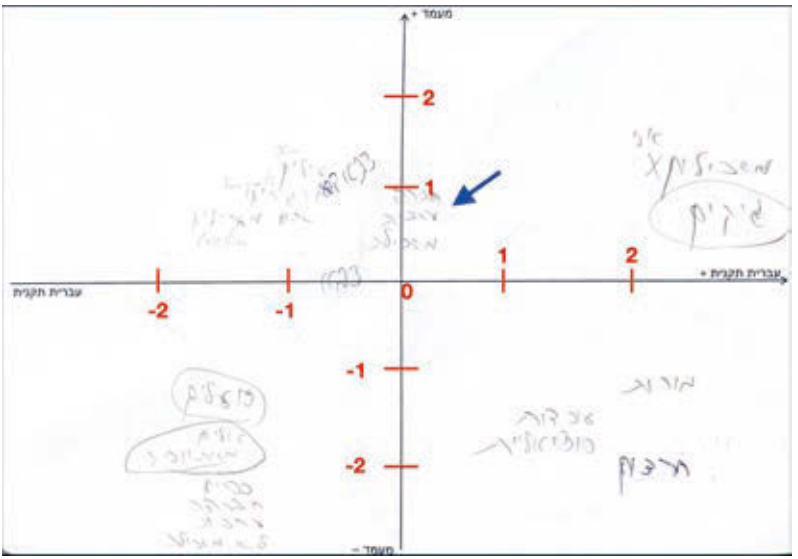

Fig. 5.4: Scale for analyzing ratings

5.5.1 Classification of the entries and summary of the ratings

After I had assigned values to all entries, I started to look for patterns in the data. Except for the recurrent categories which were defined in 5.4.3, it was hard to detect any correlations between semantic categories of the entries and their ratings. I further classified every entry, based on semantic criteria in conjunction with the ratings, to establish more general categories for their summary. Because I wanted to arrive at categories on a higher conceptual level than most of the entries' simplifications, this process of bundling depended on more theoretical premises than the process of simplification. At first, I bundled semantically similar entries into categories which could be understood as potentially distinguishable on the basis of common linguistic characteristics, in the context of Q14. According to my conception, the new categories which were created with this process should be on a basic level of categorization, in that they were organized around prototypes with typical socio-demographic and linguistic characteristics.

For example, the new category 'Arabs' contains not only the 15 simplifications "Arabs," but six simplifications "Druze," two simplifications "Christians" and the entries "Daliyat al-Karmel" and "Circassians." I bundled these entries because I assumed that the participants' representations of these categories are shaped as having an 'Arabic accent' when they speak Hebrew. Thus, the common characteristic of this category is the use of Arabic. From the context of the recordings, it was clear

that "Christians" referred to 'Arab Christians' and hence native Arabic speakers. Usually, Druze are native Arabic speakers, too, and the entry "Daliyat al-Karmel" was interpreted as referring to Druze because it designates a town near Haifa which is known for its mostly Druze population. For this analytical step, I tried to classify every entry to get the largest possible quantity for comparisons. Naturally, not all entries fit equally well into the new categories and I redid the classification several times until I felt that I could move on with the analysis. "Circassians" refers to a minority group who do not speak Arabic as L1, but for religious purposes because they are Muslims. It is the category which fits least into the category 'Arabs', but I could not find any semantically similar entries with which to form a separate category.

The classification resulted in 13 categories which are listed in Table 5.4. While eight of the categories are named after existing simplifications and, in fact, the participants' own words (in-vivo codes), some categories need further explanation. The largest category 'educated,' with 33 entries from 14 different participants, contains the recurrent categories 'academics,' 'intellectuals' and 'teachers,' skilled professions such as "lawyers," "high-tech people," "physicians" and "social workers (female)." Also s41m3l1's entry "geeks" from the above discussion was classified under 'educated' because it has similar rating values as the other entries and was described as semantically close to these by s41m3l1 (see 5.3).

'Jewish elite' is a default category and refers to well established ways of life in Israel which were expressed by the entries *yalidei ha-'arets* 'native Israelis' and *'olim l'o ḥadashim* 'not new (established) immigrants'. The category includes the simplifications "kibbutsniks," "*moshavniks*" and the geographic references to "Tel Aviv" and "Herzliya residents." Religious aspects are also contained, with "national religious" and "secular." Less central to the category 'Jewish elite' are "Americans" and "French:" these potentially well-established immigrant groups are defined as belonging to the elite because of financial and ideological aspects, but are likely to be represented with different accents in Hebrew. 'New immigrants' refers to Jewish immigrants who arrived in Israel within the last five years – in the sense of the term *'olim ḥadashim* (see 5.4.2). 'Periphery' refers to notions of geographical and social marginalization which typically intersect in Israel, as discussed in Chapter 3 and especially in 3.1.5. It contains "Africans," "development towns," "*ḳrayot*" (an agglomeration close to Haifa) "low socio-economic status," "prisoners," "settlers" and "(blue collar) workers." The entry "*'arsim*" is central to this category because it refers to the stereotype of a non-*Ashkenazi* young working-class man who lives outside of the modern urban centers (Mizrachi & Herzog 2012: 428). 'Public figures' is based on the Hebrew notion of *'ish tsibori* which is defined by Schwarzwald (2007: 75) to contain, among others, members of the Knesset, artists, journalists and radio and TV broadcasters. Interestingly, Schwarzwald (2007: 75) characterizes this group

by their common aim of getting close to the people which leads them to lower their register instead of "elevating the nation to the high language.[4]"

For each of these categories, I calculated means for 'correct Hebrew' and 'social status.' As can be seen in Table 5.4, the categories can contain multiple simplifications from the same participant: nine entries which were classified under 'army' were produced by just six participants. Therefore, I had to calculate an average value for all the simplifications from the same participant which were classified together before calculating the overall average for the category. For example, I calculated means out of s41m3l1's two entries *tzava'* and *tzava' ḳeva'*, which were both simplified as "army" (see 5.4.2), before I calculated the mean from all participants for the category 'army.' Thereby, I made sure that the data of each participant was weighted equally for the calculation of the category means (see Table 5.4). These categories and their ratings

Tab. 5.4: Categories after classification with number of entries, different participants' referrals, mean ratings for Correct Hebrew and Status

Category	Entries	Participants	CH (m)	Status (m)
Arabs	25	10	0.22	0.18
Army	9	6	-0.11	0.69
Ashkenazim	7	7	1.29	1.14
Educated	33	14	1.26	0.88
Ethiopians	9	8	-1.00	-0.75
Haredim	13	11	0.73	-0.41
Jewish elite	24	15	0.60	1.23
Mizrahim	14	9	-0.15	0.07
New immigrants	8	6	-1.08	0.25
Periphery	13	10	-0.75	-0.35
Public figures	21	9	0.90	1.22
Russians	10	10	-0.20	0.00
Uneducated	4	4	-0.25	-1.25

will be contextualized and discussed together with the participants' statements from the interviews in Chapter 6.

4 My translation from the original: מתוך רצון להתקרב לעם, הדיבור משתופף אליו במקום להעלות את העם אל הלשון הגבוהה, במקום לטפח לשון תרבות ולשמש דוגמה.

5.5.2 Comparing categories and rating variables

I plotted means of the categories from Table 5.4 to visually compare how they were rated (see Fig. 5.5). To test the hypothesis that participants represented groups of HSs who speak correctly as having a high social status, I fitted a simple linear regression model on the mean ratings.[5] I modeled the variable 'correct Hebrew' as predictor of 'social status.' The output of the simple linear regression model indicates that mean

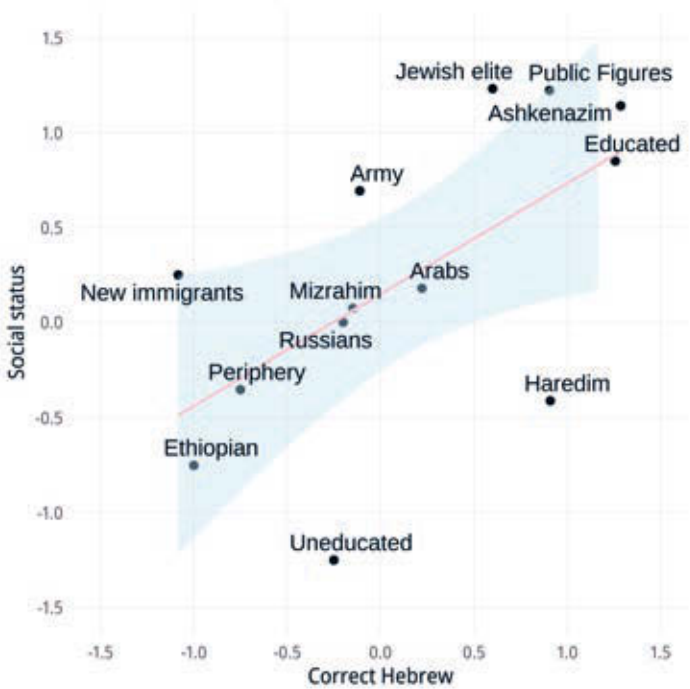

Fig. 5.5: Mean values for GERT categories with fitted linear regression model (red line) and 95% confidence interval (blue shading)

ratings for 'correct Hebrew' explained 36.8% of the variation in mean ratings for 'social status' [$F(1,11) = 6.41$, $p = .028$]. The low p-value suggests that mean ratings for 'correct Hebrew' are likely to affect mean ratings for 'social status.' The rising slope of the model, illustrated as red line, indicates that the model predicts increasing

5 I used R Statistical Software v4.3.2 (R Core Team 2023) for all analyses, and ggplot2 from the tidyverse package (Wickham et al. 2019) for creating plots.

mean ratings for ‘social status’ together with increasing mean ratings for ‘correct Hebrew.’ However, the categories outside of the shaded area somewhat contradict the predicted positive linear relation between the variables: For example, ‘Haredim’ is below the shaded area and thus outside of the confidence interval because the model would have predicted higher mean ‘social status’ based on the mean for ‘correct Hebrew.’ In the following discussion, heatmaps will be used to compare individual ratings for each category and to assess inasmuch participants diverged in their ratings – information that is obscured by aggregating means.

5.6 Typical GERT participants

The small sample size of 21 participants who completed GERT does not allow for detailed comparisons of sub-samples along socio-demographic variables. Therefore, a data-based analysis of possible patterns in the GERT corpus is more reasonable than departing from the participants’ characteristics. As noted above, there seem to be two main strategies for the completion of GERT: a minimalist approach which can be illustrated with m69f4l2’s data and a more extensive and differentiated approach just as the one represented by s41m3l1 (see Fig. 5.3). It needs to be noted that participants might have understood the task differently and used different strategies for this reason – but this does not rule out the prevalence of two different strategies. On the contrary, a certain interpretation of the task may again hint at a participant’s preference for a certain strategy.

The 21 participants who completed GERT produced 190 entries in total. On average, each participant came up with nine entries. Just one participant did not mention more than two entries, whereas the most productive participant came up with 26 entries. Interestingly, these values coincide almost exactly with the data which is presented in Lameli et al. (2008: 62) for their elicitation of “Sprachraumkonzepte” with the help of maps of Germany. Their 169 participants also produced 8.5 entries, on average, with the extremes of just two as the minimal number of entries per participant and 26 as the maximal number. These numbers strengthen the hypothesis that the number of categories used for the categorization of linguistic variation is naturally limited due to the principle of “cognitive economy” (Rosch 1978: 28–29; see 2.1.4.4). The number of categories which participants use during elicitation tasks such as GERT and mental maps seems to vary typically between four and 20.

Comparing the quantity of their entries, GERT participants can be bundled into two groups: ten participants produced seven or fewer entries and 11 participants produced between nine and 26 entries. I will refer to the group that mentioned fewer entries as ‘minimalist group’ and to the other as the ‘productive group.’ Minimalists produced only 4.3 entries on average – five of them produced exactly four entries.

Because it can be assumed that participants tried to produce exactly one entry for each of the four segments of the GERT template, it is noteworthy that only one participant completed the template in this manner. Due to the definition of the minimalist group, it can be expected that they also based their categorization on fewer conceptual domains than the productive group. Although they used only a few categories, they did not all use the same or similar ones. They also used different conceptual domains: three minimalists referred mostly to the domain of 'origin,' while the others referred mostly to 'education' and 'occupation' and to other concepts. Overall, their entries can be described as less specific. The difference in the degree of specification which is characteristic for the two groups, to which m69f4l2 and s41m3l1 belong, can be seen in Fig. 5.3.

All productive participants made use of the concept 'origin,' in addition to other concepts, which is not surprising because it is the most prominent concept among the recurrent categories (see 5.4.3). While it can be seen from s41m3l1's entries that he differentiated the categories 'Arabs,' 'army' and 'immigrants' on a finer level, it is hard to make general judgments about the degree of specificity of a category. For example, it does not seem reasonable to claim that "social workers (female)" is a more specific category than *kfarim, ḥevra 'aravia lo' maskila* 'villages, uneducated Arab society.'

It could be assumed that the minimalists can be characterized as thinking only in terms of black and white oppositions, whereas the productive group could be characterized as bean counters. Thereby, two cognitive strategies are juxtaposed: an economic, but potentially over-simplistic approach against a more precise, but cognitively costly approach which can in turn lead to ambiguous categorizations. Ultimately, these strategies are not substantiated by the GERT data – but, it would be interesting to adapt GERT methodologically to be able to investigate further into this topic.

In summary, both minimalist and productive participants used several conceptual domains at the same time and behaved inconsistently in their use of categories, as noted in 5.3. Both groups used entries which reflect the participants' personal inclinations or constraints towards the use of certain categories. For example, c36f3l1 who only produced the categories 'intellectuals' and 'uneducated' voiced concerns about political correctness – that is to say, she did not want to categorize along the concept 'origin.' The only distinguishable characteristic about the groups' entries is that the minimalists used less specific entries which is not surprising since they produced much fewer entries, overall. For example, m69f4l2's categories can be described as more basic and seem to be on a similar conceptual level, although these properties are hard to determine absolutely.