

ICS 03.080.01
CCS A 10



中国语言服务行业规范

Specifications for the Language Service Industry in China

ZYF 00x--2021

中国特色话语翻译 高端语料库建设 第3 部分：抽样检验

Translations of Chinese key terms and expressions — Corpus
construction — Part 1: Sampling inspection

20**-**-*发布

20**-**-*实施

中国翻译协会 发布

目 次

前 言.....	II
引 言.....	III
1 范围.....	1
2 规范性引用文件.....	1
3 术语和定义.....	1
4 符号和缩略语.....	2
5 基本思想.....	3
6 抽样方案的关键参数.....	3
6.1 批和批量.....	3
6.2 检验水平.....	3
6.3 接收质量限.....	3
6.4 抽样方案类型.....	4
7 抽样检验流程.....	4
7.1 需求分析.....	5
7.2 抽样参数确定.....	5
7.3 抽样方案检索.....	6
7.4 抽样检验实施.....	9
7.5 批的再提交.....	12
8 抽样风险处理.....	12
附录 A（资料性） 抽样方案检索示例.....	13
参 考 文 献.....	17

前 言

本文件按照 GB/T 1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

本文件是 ZYF 00X《中国特色话语翻译 高端语料库建设》的第3部分。ZYF 00X 已经发布了以下部分：

——第1部分：基本要求

——第2部分：系统架构

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由。。。提出。

本文件由。。。归口。

本文件起草单位：

本文件主要起草人：

引 言

讲好中国故事，传播好中国声音，展示真实、立体、全面的中国，是加强我国国际传播能力建设的重要任务。加强国际传播能力建设，有助于形成同我国综合国力和国际地位相匹配的国际话语权，为我国改革发展稳定营造有利外部舆论环境，有助于推动构建人类命运共同体。

利用信息化和语料库技术构建中国特色话语翻译语料库，有助于中国特色话语的阐释与传播，进一步拓展中国特色话语传播途径、拓宽传播方式、创新传播机制，增进国际社会对中国的理解和认同，争取更多国际话语权，提升中国软实力。中国特色话语翻译高端语料库建设系列标准旨在通过规范语料库建设过程、系统平台、质量控制方法等，提高语料库建设效率和质量，拟由三个部分组成：

——第1部分：基本要求。目的在于给出语料库建设基本原则、基本流程、步骤、管理要求。

——第2部分：系统架构。目的在于明确语料库建设系统的主要架构和功能。

——第3部分：抽样检验。目的在于给出对语料库质量合格情况进行检验评估的程序、流程、判定规则等。

中国特色话语翻译 高端语料库建设 第3部分：抽样检验

1 范围

本文件规定了利用抽样统计理论对中国特色话语翻译高端语料库(以下简称“语料库”)质量合格情况进行检验评估的基本思想、抽样方案的关键参数、抽样检验流程、抽样风险处理等。

本文件适用于在无法对语料库所有内容质量进行全面检验时对语料库建设进行质量控制、验收评价、等级评定、资源管理等工作,其他语料库建设管理工作可参照使用。

本文件不包括语料库的具体检验项目、质量要求、检验技术、检验工具等。

注1:本文件给出对语料库进行抽样检验的程序规则,不涉及具体检验的特性、要求等,抽样检验结果以“合格/不合格”“满足/不满足”“正确/不正确”“接收/不接受”等形式给出,而不是“好/坏”或“一级/二级”等。

注2:对于语料质量的具体要求(如译文是否正确、标注是否规范等)以及判别标准等,是由每个具体的语料库建设或检验任务来确定,不在本文件范围内。

注3:无法对语料库所有内容质量进行全面检验,可能是由于语料库规模太大、无法通过软件工具自动进行、或由于时间紧或成本高无法全面展开等多种原因。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中,注日期的引用文件,仅该日期对应的版本适用于本文件;不注日期的引用文件,其最新版本(包括所有的修改单)适用于本文件。

ZYF 00X—2021 中国特色话语翻译 高端语料库建设 第1部分:基本要求

GB/T 2828.1—2012 计数抽样检验程序 第1部分:按接收质量限(AQL)检索的逐批检验抽样计划

3 术语和定义

ZYF 00X—2021界定的以及下列术语和定义适用于本文件。

3.1

检验 inspection

为确定中国特色话语翻译高端语料库的各特性是否合格,测量、检查、测试语料库的一种或多种特性,并且与规定要求进行比较的活动。

[来源:GB/T 2828.1—2012, 3.1.1, 有修改]

3.2

初次检验 original inspection

按照本文件的规定对批进行的第一次检验。

[来源:GB/T 2828.1—2012, 3.1.2, 有修改]

注:对于已判定为不符合质量要求的批,经修改后再次提交检验,不属于初次检验。

3.3

批 lot

汇集在一起的一定数量的语料记录。

[来源：GB/T 2828.1—2012，3.1.13，有修改]

注：本文件的“批”不同于语料库建设提交过程中的建设和交付时用到的“批”的概念。前者表示为进行抽样而对语料库记录进行分组后每组的规模，是抽样过程中对记录的重新分组。后者表示语料库建设阶段成果规模，是按建设阶段或提交时间来划分的。因此，一次提交的数据也可能会分为多个抽样的批，一个抽样批中也可能包括多次提交的数据。

3.4

批量 lot size

批中包含的语料记录的条数。

[来源：GB/T 2828.1—2012，3.1.4，有修改]

3.5

样本 sample

取自一个批并且能提供该批信息的一个语料记录。

[来源：GB/T 2828.1—2012，3.1.15，有修改]

3.6

样本量 sample size

样本中所包含的语料记录的个数。

[来源：GB/T 2828.1—2012，3.1.16，有修改]

3.7

抽样方案 sampling plan

样本量和判定批符合质量要求准则的组合。

[来源：GB/T 2828.1—2012，3.1.17]

3.8

正常检验 normal inspection

当过程平均（见 3.1.12）优于接收质量限（见 3.1.13）时，所使用的一种能保证批以高概率判定符合质量要求的抽样方案的检验。

[来源：GB/T 2828.1—2012，3.1.20]

3.9

加严检验 tightened inspection

使用比相应正常检验抽样方案判定符合质量的准则更严厉的判定准则的一种抽样方案的检验。

[来源：GB/T 2828.1—2012，3.1.21，有修改]

3.10

过程平均 process average

在规定的时段或语料库建设内质量水平的平均。

[来源：GB/T 2828.1—2012，3.1.25]

注：在本部分中，过程平均是过程处于统计控制状态期间的质量水平的平均。

3.11

接收质量限 acceptance quality limit

当一个连续系列批被提交验收抽样时，可容忍的最差过程平均质量水平。

[来源：GB/T 2828.1—2012，3.1.26]

4 符号和缩略语

Ac: 接收数 (Accept)
AQL: 接收质量限 (Acceptance Quality Limit)
N: 批量 (lot size)
n: 样本量 (sample size)
Re: 拒收数 (Reject)

5 基本思想

本文件所给出的抽样检验方法是从待检验的语料库中随机抽取部分语料记录（即样本）进行检验，根据语料库合格率（或不合格率）要求、语料总体规模、抽样过程、样品中不合格品的数量等信息，分析判断语料库整体质量是否达到指定合格率（或不合格率）的一种方法。

注：本文件不是计算语料库实际合格率（或不合格率）到底是多少，而是判断其质量是否满足某个合格率（或不合格率）要求。其思想是，只要满足合格率要求即可判为质量合格，从而减少样本量和计算复杂度。

由于抽样的随机性，对语料库整体质量的判断会存在一定误判或漏判风险。通过合理确定抽样参数等措施可以有效降低这些风险。

6 抽样方案的关键参数

6.1 批和批量

总体而言，批和批量的确定主要和语料库建设人员专业水平、语料库建设流程、质量要求高低、语料库错误影响程度、管理工作量大小、检验时间要求、成本要求等因素有关。

总体而言，在语料规模相同的情况下，增加批次，检验工作的误判和漏判风险会下降，检验所需总样本量会增加，检验总成本、管理成本也会增加。

6.2 检验水平

检验水平反映了对抽样检验工作质量管理要求的严格程度。本文件中，检验水平从低到高依次为：水平I、水平II和水平III。

如希望漏判风险和误判风险较低、提高鉴别能力、提高质量管理水平，则需要较高的检验水平，需要的样本量也较多；如对漏判风险、误判风险、鉴别能力等要求不高，则可降低检验水平，以减少样本量。

6.3 接收质量限

本文件中，接收质量限（AQL）以不合格记录所占百分比来表示，是语料库所有方所允许的最大不合格率。本文件中，不合格率与合格率的相互关系为：不合格率=1-合格率。

每个语料库只能设置一个AQL值，语料库中的每个数据表只能设置一个AQL值。

AQL值越大，表明允许的不合格记录所占比率越大，也就是质量要求越低，对抽样方案的严格程度要求也就越低；反之，AQL值越小，表明允许的不合格记录所占比率越小，也就是质量要求越高，对抽样方案的严格程度要求也就越高。

按照GB/T 2828.1-2012中第11章所给出的判别标准，如实际抽样检验结果判别为接收，则表明语料库的实际不合格率不超过接收质量限，即不超过允许的不合格率，则通过抽样检验，认为语料库质量合格；反之则表示未通过抽样检验，语料库质量不合格。

AQL是抽样方案设计和实施抽样检验的重要指标，其设置应遵循以下原则：

- a) 应能够满足语料库建设、管理的实际需求，并兼顾考虑工作成本等要求，不宜设置过高或过低；
- b) 应考虑语料应用服务过程中所可能带来的经济和社会影响，对于影响较大的，AQL宜设置较低；
- c) 同一语料库下不同数据表的AQL设置可以不同，但AQL较高的数据表会影响语料库整体质量。

6.4 抽样方案类型

本文件中，当对语料库进行抽样和判定时，抽样方案类型主要分为两种：

- a) 一次抽样方案：一次性抽取所有样本并进行判定；
- b) 二次抽样方案：先抽取一定数量样本并进行判定，根据判定结果确定是否再从该批剩余记录中抽取一定数量的样本进行第二次检验和判定。

在其他各项参数和要求都相同的条件下，不同抽样方案类型达到的效果几乎相同。对同一语料库进行抽样检验时，不同抽样方案类型间不宜混用。

不同抽样方案类型各有特点，简单对比如表1所示：

表1 不同抽样方案类型比较

特点比较	方案类型	
	一次抽样方案	二次抽样方案
每批平均样本量	多	少
每批样本量变动	无	有
管理难度	低	高
操作难度	低	高
操作性	简单	复杂

注：从管理角度，总体来说，一次抽样简单，对批判断的失误少。二次抽样管理较复杂，演算及记录比较困难、繁琐，对检验人员的能力和知识要求较高。

当语料库整体的质量情况预估很好或者很差时，可选用二次抽样方案。

7 抽样检验流程

中国特色话语翻译高端语料库抽样检验的主要流程如图1所示：

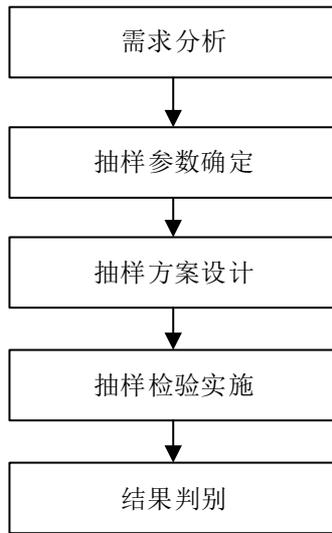


图 1 抽样检验流程图

7.1 需求分析

结合具体语料库内容、检验要求、建设要求、应用需求、管理成本等信息，从以下方面综合分析语料库抽样检验工作具体要求：

- a) 质量水平要求；
- b) 语料库规模；
- c) 语料库结构；
- d) 语料库建设流程；
- e) 单个样本检验工作难度；
- f) 检验工作误判和漏判可能带来的风险和危害；
- g) 经费、时间、以及其他检验工作所依赖的资源限制。

7.2 抽样参数确定

7.2.1 批和批量的确定

根据ZYF 00X—2021，语料库通常由多个表组成，每个表存储不同类型数据，不同表的结构可能不同，每个表内存储多条语料数据。当语料库规模较大或持续建设时，可按照下列步骤将语料库分解为多个批之后进行分批检验，如图2所示。

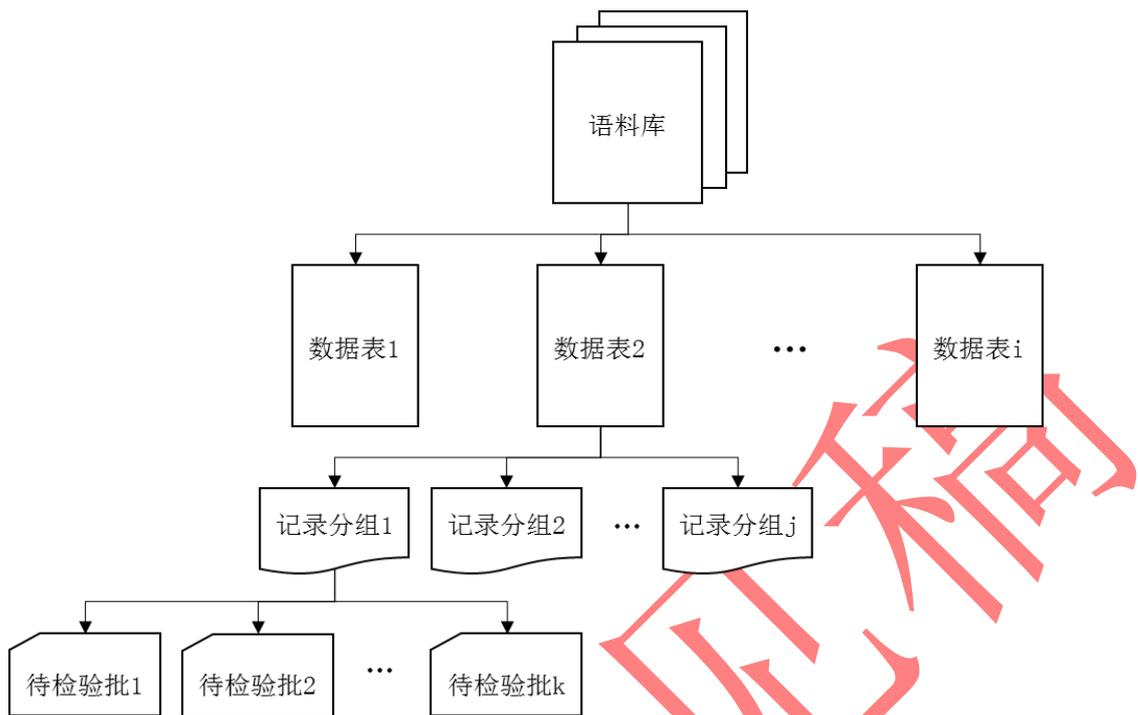


图2 语料库分解为待检验批示意图

基本分批要求如下：

- a) 对语料库中每个数据表，对表中的语料记录进行分组，分组时应满足以下要求：
 - 1) 语料库加工过程相同、参与人员（主要是翻译和标注人员）相同、机器翻译软件版本、以及自动标注工具版本都相同，可分为一组；
 - 2) 语料库加工过程、参与人员（主要是翻译和标注人员）、机器翻译软件、自动标注工具中任何一个不不同时，不宜分为一组；
- b) 对每个记录分组根据管理需求和6.1进行分批。

7.2.2 检验水平的确定

在没有特殊要求和规定时，宜使用检验水平 II。当有必要提高鉴别力时，应选择检验水平 III。当必须减少样本量时，可选择检验水平 I。

7.2.3 接收质量限 AQL 的确定

本文件中，如无特殊要求，可将相关质量要求规定中的不合格率视为 AQL 值，也可根据需要确定一个更严格的 AQL 值。

如果给出的是语料库整体的 AQL 值，那么每个数据表的 AQL 值应不大于语料库的 AQL 值，每个数据表可根据其重要性、影响程度等设置不同的 AQL 值。如果是给出的每个数据表的 AQL 值，语料库的总体 AQL 值宜选择所有数据表的最大 AQL 值。

注 1：相关质量要求规定一般会在语料库建设合同、任务书等文件中给出。

注 2：根据 GB/T2828.1，AQL 值一般不超过 10%。

7.3 抽样方案检索

7.3.1 AQL 值可直接检索到

当已确定的 AQL 值可以直接在 GB/T 2828.1—2012 表 2 和表 3 中检索到时，根据检验

工作要求和确定的参数，按照 GB/T 2828.1—2012 中 10.3 的要求，检索确定抽样方案。示例见附录 A。

7.3.2 AQL 值不可直接检索到

当已确定的 AQL 值不能直接在 GB/T 2828.1—2012 表 2 和表 3 中检索到时，可按照 7.3.1 的要求，从表中按与原 AQL 相近的 AQL 进行检索，也可按照，可根据图 3 和表 2 计算求得抽样方案。

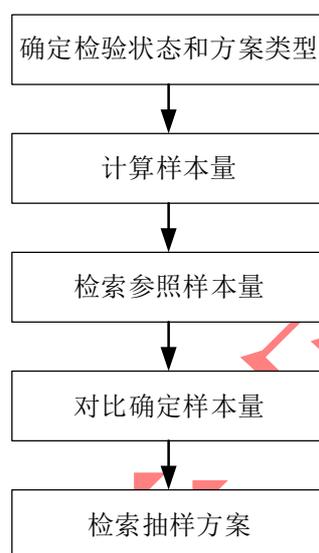


图 3 利用 AQL 计算抽样方案流程

表2 抽样方案计算汇总表

方案类型	样本量系数	累计样本量系数	接收质量限 (AQL) — 正常检验											
			12.5/n	50/n	80/n	125/n	200/n	315/n	—	500/n	—	800/n	—	1250/n
			Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re
一次	1	1	0 1	1 2	2 3	3 4	5 6	7 8	8 9	10 11	12 13	14 15	18 19	21 22
二次	0.63	0.63	*	0 2	0 3	1 3	2 5	3 6	4 7	5 9	6 10	7 11	9 14	11 16
	0.63	1.26		1 2	3 4	4 5	6 7	9 10	10 11	12 13	15 16	18 19	23 24	26 27
			20/n	80/n	125/n	200/n	315/n	—	500/n	—	800/n	—	1250/n	—
			接收质量 (AQL) — 加严检验											
注: *表示采用一次抽样方案														

具体计算流程如下。

- a) 确定检验状态和方案类型：确认需要检索方案的检验状态为正常检验还是加严检验，抽样方案类型为一次抽样方案还是二次抽样方案。
- b) 计算样本量：根据表 2 计算样本量。如果是正常检验，则根据表 2 第一行数值计算，若为加严检验，则根据表 2 最后一行数值计算。具体计算方法如下：
 - 1) 正常检验，表 2 第一行数值即 $12.5/n$ 、 $50/n$ 、 $80/n$ …… $1250/n$ 为正常检验时的 AQL 值，在 AQL 值确定后样本量 n 通过计算可得，即 $n=12.5/AQL$ ， $n=50/AQL$ ，……， $n=1250/AQL$ ，据此依次计算出第一行的 n 值；
 - 2) 加严检验，表 2 最后一行数值即 $20/n$ 、 $80/n$ 、 $1250/n$ …… $1250/n$ 为加严检验时的 AQL 值，在 AQL 值确定后样本量 n 通过计算可得，即 $n=20/AQL$ ， $n=80/AQL$ ，……， $n=1250/AQL$ ，据此依次计算出最后一行的 n 值。若为一次抽样方案，样本量系数为 1，计算所得的 n 值即为样本量值；若为二次抽样方案，因其样本量系数为 0.63，计算所得的 n 值乘以 0.63 即为二次抽样方案的单次样本量。
- c) 检索参照样本量：根据已确定批量与检验水平通过 GB/T 2828.1—2012 表 1 检索相应的样本量字码，再根据样本量字码对照 GB/T 2828.1—2012 表 2A、表 2B、表 3A、表 3B 查找相应检验状态和方案类型下样本量字码对应的样本量，将此样本量作为参照样本量。
- d) 对比确定样本量：将参照样本量与 b) 中计算所得的样本量进行比较，选定 b) 中最接近参照样本量的值。
- e) 检索抽样方案：检索表 2，根据 d) 中选定的样本量所在列，结合方案类型所在行，二者交叉处即为抽样方案的接收数与拒收数，即可获得抽样方案。

7.4 抽样检验实施

7.4.1 基本方法

除对检验有特殊要求之外，检验开始时一般采用正常检验。

7.4.2 确定抽取样本的方法

样本抽取采用分层抽样的方法。按语料库中的数据表进行分层，然后按比例从不同层中（即数据表中）使用简单随机的方式抽取样本。

简单随机抽样，即通过随机数表等方法产生随机数，对应到语料记录，抽取对应语料记录作为样本。

7.4.3 每批结果判别

根据每批样本的不合格记录的数量，判断该批是否通过检验，通过检验则称为“接收”，未通过检验则称为“不接收”。

不接收的批的处置按照 GB/T 2828.1—2012 第 7 章规定执行。

注：本文件中的“接收”、“不接收”仅表示批的抽样检验是否合格，是为了与 GB/T 2828.1—2012 保持一致。实际应用中，对于“接收”的批中语料记录能否进入下一个业务环节，需要根据实际业务场景和管理要求来确定，不在本文件范围内。

样品质量和正确性的分析判断工作可通过机器自动、人机交互、专家评判等形式结合进行。

当专家评判时，应由中高级以上资质译员负责实施。

评判工作主要根据以下要求：

-
- a) ZYF 00X—2021 中的规定的审校要求；
 - b) 语料库项目任务书或合同中规定的要求；
 - c) 项目委托方和建设方共同确认的其他要求。

7.4.4 转移规则及结果判别

宜根据图 4 给定的规则，对连续各批进行检验和判别。

征求意见稿

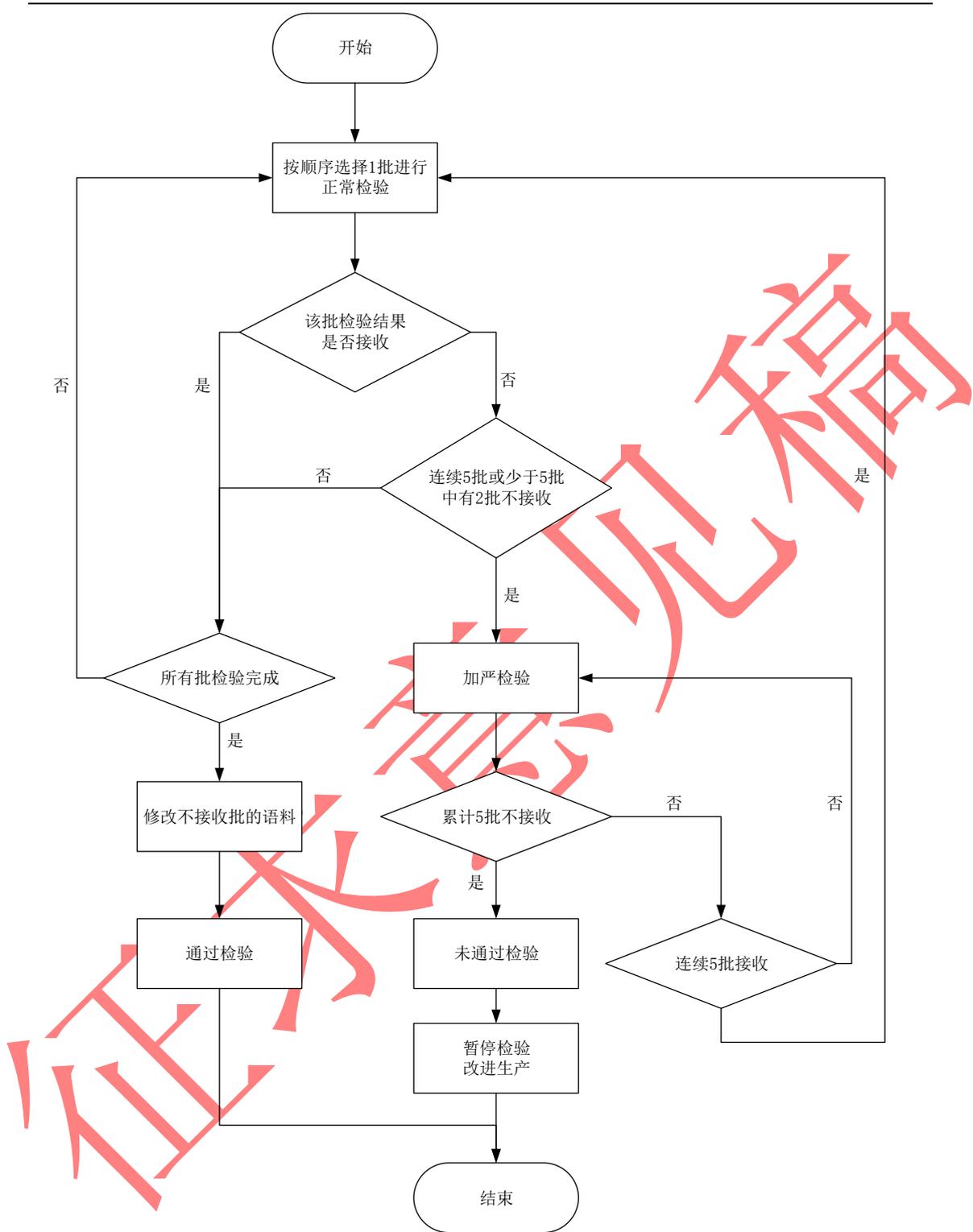


图4 转移规则

加严检验时，应按照 GB/T 2828.1—2012 表 2-B、表 3-B 进行检索抽样方案并执行相应的检验。

注：本文件不使用放宽检验。

7.4.5 暂停检验及改进生产

如因不接收批次数量累计达到上限，应暂停检验，应对语料库建设过程进行全面检查，分析不合格记录产生原因，改进语料库建设工作。

7.5 批的再提交

对于拒收批，如果可能，应对整批进行筛选，剔除所有不合格记录，并对不合格记录加以修改或者替换（剔除的不合格记录经过修改校正后可选择归入原批），再次组批交验，即再次提交。再次提交时，对导致拒收的这类不合格都应进行检验，因重新修改或替换可能引起的其他类不合格也应进行检验，除此之外的其他类别的项是否进行检验可根据实际需要而定。

8 抽样风险处理

抽样检验存在一定的错误判定风险，即抽样风险。抽样风险是由于抽取的样本有限而不能真实反映语料库的质量所造成。抽样风险是源自抽样本身的风险，无法完全避免。

抽样风险主要分为误判风险和漏判风险：达到语料库质量要求但被判定为不合格的为误判，体现为误判风险；未达到语料库质量要求但被判定为合格的为漏判，体现为漏判风险。抽样风险受到样本量、判定数以及检验水平等影响。

可通过以下方法降低误判风险：

- a) 提高改进对样本进行检验的技术、方法、流程等；
- b) 其他参数不变时，提高检验水平；
- c) 对检验人员进行专业培训，提高专业能力，必要时应由专家参与检验；
- d) 进行复检复查。

可通过以下方法降低漏判风险：

- a) 其他参数不变时，提高检验水平；
- b) 其他参数不变时，降低批量，增加批次；
- c) 利用软件系统或工具辅助检验；
- d) 进行复检复查。

附录 A
(资料性)
抽样方案检索示例

A.1 AQL值可直接检索到

对某语料库进行连续分批的抽样检验，根据实际情况确定的抽样参数为：批量为5000，采用检验水平II，AQL值为0.1，使用一次抽样方案进行检验，抽样方案检索步骤如下：

初次检验时使用正常检验，确定的抽样方案为一次抽样。

- a) 首先检索样本量字码，依据GB/T2828.1-2012表1，根据批量5000所在的行和检验水平II所在的列找到交叉位置的样本量字码为L（如图A.1所示）；
- b) 检索正常检验的抽样方案，依据GB/T2828.1-2012表2-A，根据样本量字码L所在行和AQL值0.1所在列交叉位置寻找抽样方案，查表可知此交叉位置为向上箭头（如图A.2所示），根据标准，选择箭头所指向的上—行的方案作为抽样方案，这里确定的抽样方案为（125,0,1）；
- c) 按照抽样方案（125,0,1），抽取125个样本进行检验，接收数为0，拒收数为1，即如果未发现不合格则该批检验通过，如果发现1项检验不合格则该批不能通过检验。

在连续批的检验中，如果出现连续5批或少于5批中有2批不接收，则实施加严检验，加严检验的抽样方案依据GB/T 2828.1-2012表2-B进行检索，检索方法同正常检验。

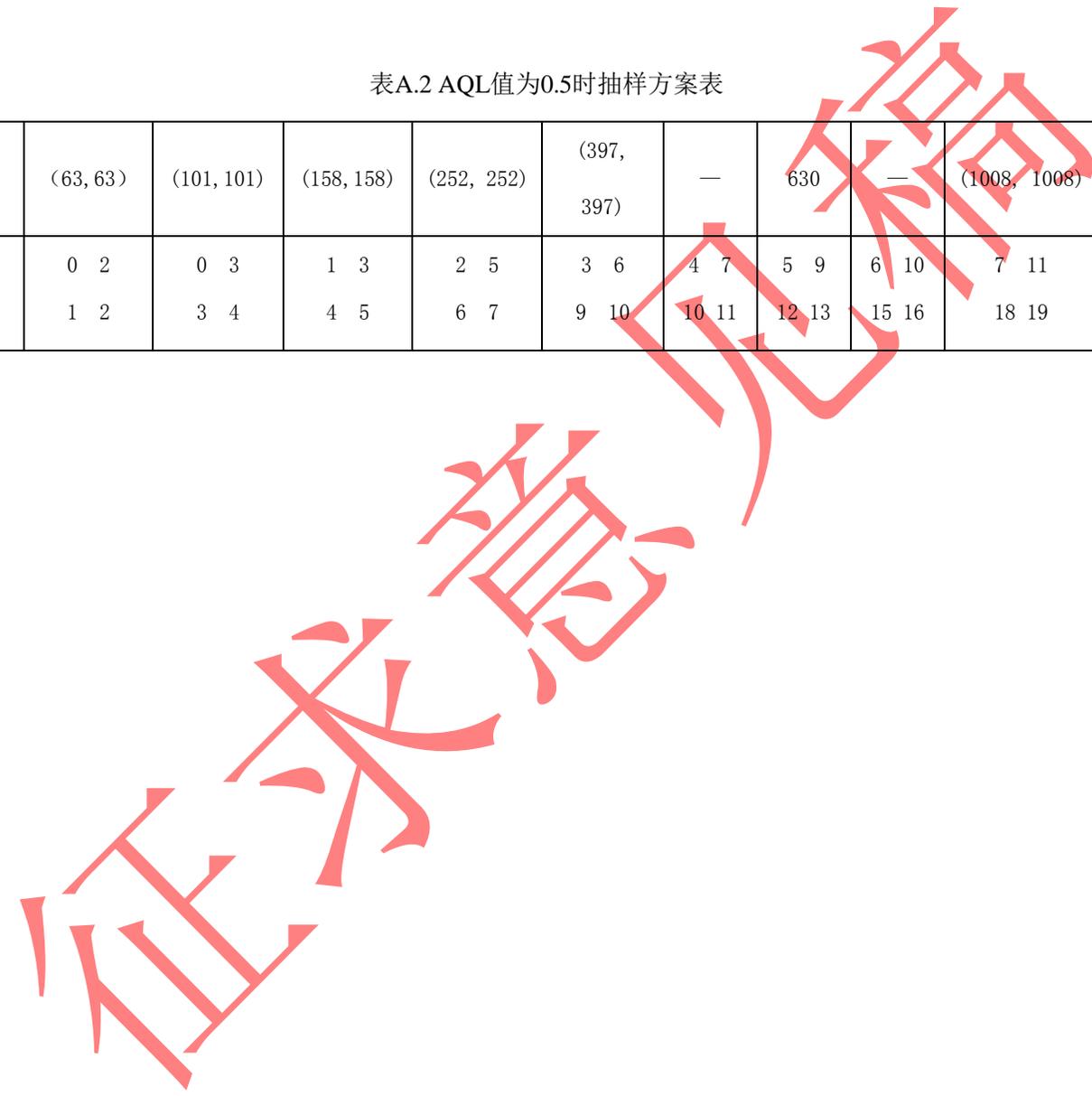
表 1 样本量字码(见 10.1 和 10.2)

批 量	特殊检验水平				一般检验水平		
	S-1	S-2	S-3	S-4	I	II	III
2~8	A	A	A	A	A	A	B
9~15	A	A	A	A	A	B	C
16~25	A	A	B	B	B	C	D
26~50	A	B	B	C	C	D	E
51~90	B	B	C	C	C	E	F
91~150	B	B	C	D	D	F	G
151~280	B	C	D	E	E	G	II
281~500	B	C	D	E	F	H	J
501~1 200	C	C	E	F	G	J	K
1 201~3 200	C	D	E	G	H	K	L
3 201~10 000	C	D	F	G	J	L	M
10 001~35 000	C	D	F	H	K	M	N
35 001~150 000	D	E	G	J	L	N	P
150 001~500 000	D	E	G	J	M	P	Q
500 001 及以上	D	E	H	K	N	Q	R

图A.1 样本量字码检索示意图

表A.2 AQL值为0.5时抽样方案表

样本量组 (n_1, n_2)	(16,16)	(63, 63)	(101, 101)	(158, 158)	(252, 252)	(397, 397)	—	630	—	(1008, 1008)	—	(1575, 1575)
Ac Re	*	0 2	0 3	1 3	2 5	3 6	4 7	5 9	6 10	7 11	9 14	11 16
二次抽样		1 2	3 4	4 5	6 7	9 10	10 11	12 13	15 16	18 19	23 24	26 27



-
- b) 根据批量3000，检验水平II，检索GB/T 2828.1—2012表1得到样本量字码为K，再根据GB/T 2828.1—2012表3-A，可知样本量字码为K对应的样本量为(80,80)，把(80,80)作为参照样本量；
- c) 从表A.2中可知，样本量 (63,63)与参照样本量 (80,80)最接近，因此选定样本量(63,63)。
- d) 根据标准正文中表2，样本量(63,63)对应的列与二次抽样方案对应的行交叉处的接收数和拒收数分别为0，2和1，2，因此抽样方案为(63,63|0, 2;1,2)。

加严抽样检验方案的获得方法与正常方案类似，计算样本量时要选择标准正文中表2最后一列值进行计算。

参 考 文 献

- [1] GB/T 2828.10 计数抽样检验程序 第 10 部分：计数抽样系统导则
- [2] GB/T 2828.2-2008 按极限质量（LQ）检索的孤立批检验抽样方案
- [3] GB/T 2828.11-2008 计数抽样检验程序 第 11 部分：小总体的声称质量水平的评价程序
- [3] GB/T 6378.1-2008 计量抽样检验程序 第 1 部分：按接收质量限（AQL）检索的对单一质量特性和单个 AQL 的逐批检验的一次抽样方案
- [4] GB/T19000-2008 质量管理体系 基础和术语
- [5] 张玉柱,丁文兴.《计数抽样检验程序 第 1 部分：按接收质量限（AQL）检索的逐批检验抽样计划》理解与实施[M],2012,中国质检出版社.